REPORT RESUMES

ED 003 660 56

STUDIES OF BYPASSING AS A WAY OF ADAPTING SELF-INSTRUCTION PROGRAMS TO INDIVIDUAL DIFFERENCES. FINAL REPORT.

BY- CAMPBELL, VINCENT N. BRIGGS, LESLIE J.

AMERICAN INST. FOR RESEARCH IN BEHAVIORAL SCIENCES

REPORT NUMBER NDEA-VIIA-520 PUB DATE MAY 62

REPORT NUMBER AIR-C41-5-62-FR

GRANT OEG-736103-00

EDRS PRICE MF-$0.18 HC-$3.24 81P.

DESCRIPTORS- *AUTOINSTRUCTIONAL PROGRAMS, *PROGRAMED INSTRUCTION, *INDIVIDUAL DIFFERENCES, INSTRUCTIONAL MATERIALS, LEARNING PROCESSES, LEARNING ACTIVITIES, SAN MATEO, CALIFORNIA

POTENTIAL ECONOMIC FORMS OF ADAPTIVE, SELF-INSTRUCTIONAL PROGRAMS WHICH WOULD BE OF PRACTICAL USE IN ADJUSTING TO INDIVIDUAL DIFFERENCES WERE INVESTIGATED. "PARALLEL" PROGRAMS (ONE STUDENT LEARNING BEST BY ONE PROGRAM AND A SECOND STUDENT BENEFITING MOST FROM ANOTHER) WERE FIRST EXPLORED AS A WAY TO ADJUST TO INDIVIDUAL DIFFERENCES. THIS APPROACH WAS FOUND TO BE INADEQUATE, POSSIBLY BECAUSE OF THE CHOICE OF INAPPROPRIATE KINDS OF DIFFERENCES AMONG PROGRAMS EXPECTED TO INTERACT WITH INDIVIDUAL STYLES OF LEARNING ACTIVITY. THE SECOND APPROACH TURNED TO THE DEVELOPMENT OF A SERIES OF 10 EXPERIMENTS IN THE BYPASSING FORM OF BRANCHING PROGRAMS. ABOUT 2,000 PUBLIC SCHOOL STUDENTS SERVED AS SUBJECTS. MANY WERE OBSERVED SINGLY IN LABORATORY EFFORTS TO FIND CLUES FOR IMPROVING THE BYPASS TECHNIQUE BY BOTH EXPERIMENTAL AND OBSERVATION METHODS. THESE EXPERIMENTS DID NOT ACHIEVE THE GOAL OF DEVELOPING A SIMPLE AND EFFECTIVE FORM OF ADAPTIVE PROGRAM FOR THE KINDS OF MATERIAL EMPLOYED. THE RESEARCH DID RESULT, HOWEVER, IN DEVELOPMENT OF AN EMPIRICAL METHOD FOR DETERMINING THE DEGREE TO WHICH GIVEN MATERIAL HAS THE CHARACTERISTICS FOR WHICH BYPASSING PROGRAMS SHOULD BE MOST EFFECTIVE. IN ADDITION, PROGRAMS AND GRADE NORMS FOR A WIDE RANGE OF AGE GROUPS WERE ESTABLISHED. (JH)

ED003660

520 520

AIR-C41-5/62-FR

STUDIES OF BYPASSING AS A WAY OF ADAPTING SELF-INSTRUCTION PROGRAMS TO INDIVIDUAL DIFFERENCES

Vincent N. Campbell

May 1962

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

U. S. Office of Education
Grant No.: Title VII 736103.00

American
Institute
for
Research

AMERICAN INSTITUTE for RESEARCH
PENINSULA OFFICE

8 West 41st Avenue, San Mateo California . . . Fireside 1-9297

STUDIES OF BYPASSING AS A WAY OF ADAPTING SELF-INSTRUCTION PROGRAMS TO INDIVIDUAL DIFFERENCES -FINAL REPORT

Vincent N. Campbell

AIR-C41-5/62-FR

Final Report
Grant No.: Title VII 736103.00
Leslie J. Briggs, Principal Investigator

May 1962

U. S. Office of Education

Department of Health, Education, and Welfare

Main Office 410 Amberson Avenue, Pittsburgh 32, Pennsylvania

ACKNOWLEDGMENTS

The research reported here could not have been accomplished without the excellent cooperation and assistance of the participating public schools in the Peninsula Area of California. We are greatly indebted to the superintendents, principals, staff members, teachers, and students whose efforts made possible the gathering of the data reported here.

The participating school districts are: the Hillsborough Elementary School District, Dr. Carlisle H. Kramer, Superintendent; the Menlo Park Elementary School District, Mr. Franklyn O. White, Superintendent; the San Mateo Elementary School District, Dr. John D. Andes, Superintendent; the San Mateo Union High School District, Mr. T. F. Reynolds, Superintendent; and the Sequoia Union High School District, Dr. Rex Turner, Superintendent.

Special thanks are due Donald F. Terry, Dr. Mary B. Willis, James W. Shearer, and Evelyn Wentworth for substantial help in program writing and the collection and analysis of data.

FOREWORD

In the training of teachers, few topics receive more emphasis than "individual differences." Skills to be learned by the teacher in training have to do with (a) recognition and measurement of individual differences, and (b) adjustment of instructional procedures to individual differences.

Broadly speaking, the total range of observed differences among students are dealt with by education in a variety of ways. Some differences are simply accepted as marks of the uniqueness of the individual, and are a matter of interest to the teacher as characteristics of the behavior of the person. No particular administrative or instructional adjustments are made because of these differences. Other differences become the basis for academic and vocational guidance efforts, or for assignment to particular schools, grade level, or classroom grouping.

But of most immediate concern to the teacher, once the class composition has been determined, are those differences which require the teacher to either spend much time in individual work with the student or to make special modifications in group instruction. But the teacher does not have enough time to apply a tutorial method with each student individually, and effective modifications to group instructional methods for the benefit of individuals cannot fully be made without retarding the progress of other students. Thus fully individualized instruction under normal conditions is a practical impossibility with present teacher-pupil ratios. This is especially so if we assume that maximally effective individualized instruction would require (a) varying the content of information inputs to the student; (b) varying the sensory mode of stimulation in presenting the input; (c) varying the sequences of material; (d) varying the amounts of drill, illustration, example, etc.; and (e) varying the types of student response and student activity.

If we assume that all the above kinds of variation are needed for individualized instruction, we face not only the time limitations of the teacher for individualizing instruction; we face also our lack of precise knowledge of just how the above variations in procedure can be matched to the individual. Clearly, then, special tools are needed which would enable the teacher to more effectively adapt instruction to the individual. Although we doubt that any one tool will solve the whole problem, it is this need for special tools which forms the background for the research undertaken in this project.

Programed instruction as currently conceived presents two divergent views of the whole matter of individualizing instruction. Linear programs of the most prevalent form do not concern themselves greatly about individual differences. They assume that if programs are effectively prepared for each lesson, there is potentially one sequence which will serve for all students, although the sequence will require more time for some students than for others. There is not much concern that the number of steps presented may be more than some students need. Although boredom is recognized as a possible problem in small step programs for the bright student, it is assumed that the additional instruction will do no harm. Since the typical linear programs attempt to set up conditions whereby most responses are correct during learning, and since it is assumed that this assures effective learning when well-sequenced frames also provide reinforcement, no requirement is seen for a "branching" program. For the purpose of this discussion, we will refer to these linear programs as "non-adaptive" to individual differences.

Proponents of present "branching" programs view them as a way to adjust to individual differences. When errors are made, the nature of the error is employed to determine which remedial loop of instruction the student needs. Thus branching programs assume that students may begin the instruction with varying amounts of capability for the topic, and that branching aids those who have less prior knowledge or ability to learn from the basic steps presented. Due to the use of multiple-remedial steps for each basic step, in which each remedial step is intended to correct a different misconception of the basic step, present forms of adaptive programs become quite bulky and may require either complex manipulative procedures (as in a scrambled book) or a complicated apparatus.

In the planning of the present project, we proposed to seek to develop economical forms of adaptive programs which would be of practical usefulness in adjusting to individual differences. We did not propose that such programs would be flexible in all the respects which might conceivably adjust best to such differences. That is, we sought a simple and potentially widely applicable type of adaptive program. Although our preference is to believe that composite methods of instruction are preferable to a single method, such as programed learning, for this project we restricted ourselves to convenient ways to adjust to individual differences by programed methods. (Another project, carried on under U. S. Office of Education Grant No.: Title VII 736134.01, explores the usefulness of combined methods.)

We began the project work by assuming that, in approaching a given lesson, individuals differ in their best style of performing, in such a fashion that Student A would learn best by Program A while Student B benefits most by Program B. We thus first explored the idea of using "parallel" programs as a way to adjust to these individual differences. We chose to explore this "parallel program" possibility through the first two experiments, one of which varied directness of cueing and prompting, the other varying mode of responding to the programs. As previously reported (Campbell, 1961), these two experiments failed to yield results which would encourage further effort. This could be because we chose inappropriate kinds of differences among programs expected to interact with individual styles of learning activity. Although we do not feel that further research in the parallel program approach should be discouraged, we do feel that much empirical observation of the learning activity of many individual subjects would be needed before we could choose more successful program variations expected to adjust well to individual differences. For the sake of economy, then, we chose to abandon the exploration of the parallel program approach, and turned instead to the development of a series of ten experiments in the bypassing form of branching program.

This report concentrates on eight of the ten bypassing experiments, the first two having been described in the interim report by Campbell, mentioned above. These eight experiments represent a series of attempts to concentrate effort upon first one and then another possible avenue for the improvement of the bypass type of program. In all, about 2,000 public school students served as subjects in these experiments, and several revisions and modifications of two different programs were made as the work progressed. In addition, many students were observed singly in our laboratory efforts to find clues for the improvement of the bypass technique by both experimental and observational methods.

Part One of this report presents a summary of the findings from the eight bypass experiments, and describes the rationale for believing that some of the efforts in program revision would result in effective adjustments to individual differences. For the sake of brevity and coherence, Part One omits details such as the number of subjects for each experiment, etc. These and other details are given in Part Two, experiment by experiment.

This intensive series of experimental and program revision efforts did not achieve the goal of developing a simple and effective form of adaptive program, for the kinds of material employed. However, the research did result in development of a statistical-empirical method for determining the degree to which given material has the characteristics for which bypassing programs should be most effective. The last section of Part One of this report reviews this issue in detail. From the point of view of information gained by the research, we have the facts concerning bypassing for the materials employed, and we have the programs and the grade norms for a wide age range of groups. We also have a sober appreciation of how difficult it is to develop a simple effective form of adaptive program.

Negative findings, to most experimenters, are a disappointment. We believe, however, that our efforts have value, and we suggest that similar investigations should be made of the effectiveness of the now-traditional form of branching program, which is more expensive to develop than our simpler bypass programs. Such additional research may lead to the development of more effective adaptive programs.

Were research similar to oursto be conducted with other forms of branching programs with similar findings, it would appear that the art of programing may not soon reach the goal of being adaptive to individual differences. For the immediate future, programing may succeed best through effective sequencing of linear materials, but with full use of available research regarding size of step, form of response, frequency of response, etc.

Apparently needed, then, for the learning of new materials are further efforts to develop effective adaptive programs and also efforts to improve upon present forms of linear programs. Our short linear programs did achieve some degree of learning, and the efficiency of these as compared to other linear forms needs to be known. It could turn out to be most economical to employ improved forms of linear programing for initial presentation, depending upon the teacher to function as the "remedial loop," at least until computer-controlled programing can implement effectively a greater number of the factors which should be flexibly employed to develop optimum adaptive programs.

Leslie J. Briggs
Principal Investigator

Table of Contents

Section Page

PART ONE: Bypassing--The Research Issues and Main Findings. . . . 1
- The Rationale for Bypassing 1
- Research with the Sets Program. 3
  - Overall Effectiveness of Bypassing and Remedial Loops. . 3
  - Mathematical Ability of the Learner. 14
  - The Evaluation Part of Bypassing 15
    - Validity of the Bypassing Criterion 15
    - Frequency of Evaluation 20
  - Classroom Effects on Learning. 22
- Research with the U. N. Program 24
  - Effectiveness of Bypassing and Remedial Loops. 24
- Topic Structure and Learning Objectives 29
  - Hierarchical Structure 29
  - Difficulty Interdependence 29

PART TWO: Record of Procedures, Analyses, and Results 32
- Experiment I 33
- Experiment II 35
- Experiment III 38
- Experiment IV 43
- Experiment V 45
- Experiment VI 48
- Experiment VII 51
- Experiment VIII 53

REFERENCES 62

APPENDICES 63
- A: Criterion Test for Third Sets Program
- B: Index of Difficulty Interdependence

PART ONE

Bypassing--The Research Issues and Main Findings

Encountered casually, individual differences in personality and behavior are the spice of life. But to the educator or psychologist who confronts them on duty, they are a force to be reckoned with. In trying to cause the same change in many persons, or a difference between groups, he often finds that individual differences remain the largest source of variance in the outcome. Nowhere has this been more evident than in schoolroom learning. Ability testing, now almost universal in schools, is based on the premise that these individual differences will persist.

The present research is also based on this premise. Our purpose has been to ascertain whether programed self-instruction might be improved by varying the instructional sequence so as to adjust to individual differences among learners.

Our general research strategy and early experiments are described in detail in a previous report (Campbell, 1961). Briefly, a pair of pilot experiments failed to demonstrate the practicality of employing parallel linear programs as a means for adjusting to individual differences, when variations in directness of cueing and mode of responding are the distinguishing features of parallel programs. Where only slight gains in learning efficiency are obtained, parallel linear programs hardly seem to justify their relatively great expense.

The Rationale for Bypassing

A reasonable alternative to parallel linear programs is some form of bypassing*, that is, a branching program in which everyone is presented the minimum set of basic steps, and individuals who fail a given step immediately take a remedial "loop" of instruction on that step. One bypassing program may thus offer a large number of different sequences of instruction. In principle, each learner receives a sequence which emphasizes his particular weak spots and minimizes time spent in places where he excels through acumen or prior knowledge.

---

* The report cited above discusses and compares the rationales for bypassing vs. parallel forms of branching programs, and describes our first two studies of bypassing.

A moment's thought will make it apparent that bypassing, or "adaptive instruction," is not unique to programing. The same paradigm underlies all educational plans in which instruction is intended to be geared to the learner's progress. Aside from careful attention to sequencing of material and the fact that the instructor's help is not needed to adjust the procedure to the student's progress, the main feature distinguishing self-instruction programs is the greater frequency of the learner's responses through which his progress can be evaluated. The study of bypassing programs lends new perspective to the larger and older problem of remedial instruction in general. Remedial instruction is intended to be adaptive to the learner's progress, by definition. But incisive studies of its two main properties, validity of evaluation and effectiveness of remedial learning, seem peculiarly rare in the literature (Curr & Gaulay, 1953). Perhaps this is because it seems obvious that extra instruction where the learner has failed will be beneficial. We set out to prove this truism in the realm of programed instruction and were taken aback by the findings, which are reported in the succeeding pages.

On *a priori* grounds, the appropriateness of bypassing for a given learning topic depends on the extent of individual differences in amount of instruction needed. Given an appropriate topic, the success of bypassing depends mainly on two things: the validity of the decision which directs the learner either to begin a remedial loop or to bypass it, and the effectiveness of the remedial instruction contained in the loops. In other words, one must accurately evaluate the learner's progress and also know what to do if his learning is inadequate. In this framework, the following aspects of adaptive instruction emerge as primary targets for research:

(1) Regarding evaluation of learning progress--
  (a) How should the learner's progress be evaluated? Which methods maximize predictive validity and minimize cost and time?

(b) How often should learning be evaluated for branching decisions?

(2) Effectiveness of the remedial instruction is the substance of bypassing. For those who fail a step initially, what remedial instruction will succeed? This appears to be the most demanding problem and most of our research efforts were directed to it.

(3) The other major consideration is the nature of the learning task or subject matter. It seems reasonable that bypassing would hold more potential for some topics and learning objectives than for others.

We confined our research on bypassing to programed self-instruction, but the above research questions are stated in general form because we believe our research findings with bypass programs are relevant to other instructional methods, particularly within the broad class of printed-matter presentations. In view of the mixed and generally unimpressive research findings on the effectiveness of particular features of programed instruction (Lumsdaine & Glaser, 1960; Briggs et al., 1962) it seems wise to blend programing into a larger framework of educational methods. This may help to disseminate the useful ideas generated by programers, and to remove gently the program-as-panacea atmosphere which might otherwise lead to reaction and total rejection of programing concepts.

Research with the Sets Program

Overall Effectiveness of Bypassing and Remedial Loops

The mainstream of our research consisted of evaluating the composite effectiveness of bypassing with successively revised versions of a self-instruction program on the topic of set theory. A parallel but smaller line of research was conducted with a program on the structure and functions of the United Nations. Neither program in any version represented a particular school of programing. Programed texts or booklets were the mode of presentation. Responses of various kinds were required of the learner irregularly and not too frequently. Previous evidence (e.g. Goldbeck & Campbell, 1962) had suggested that frequent responding is inefficient for comprehension learning. The correct answers appeared on the back of the same page in the Sets programs, and the United Nations (U.N.) program used similar means of providing immediate feedback.

We tried to give as much care to writing the basic steps required of all learners as to the remedial loops. Finding that a well-written loop teaches more than a poorly written basic step would be of little interest. Our main writing guide, beyond having the topic well in mind, was to keep it as clear and interesting as we could and in the simplest language possible. We tried to keep the basic steps concise in order to maximize the time that might be saved through bypassing. The remedial loops changed substantially with successive revision, in ways to be described.

Fundamentally the same research design served throughout in assessing the effectiveness of bypassing and the remedial loops. Students were assigned randomly to one of the following learning conditions:

> BYPASS: This was the experimental method already described generally. After each basic step the student was directed to skip to the next basic step if his answer was right, otherwise to continue on from where he was (through the remedial loop of instruction).
>
> SHORT (linear): In this learning method the remedial loops were omitted from the program. The student worked through the series of basic steps in linear fashion.
>
> LONG (linear): In this method the student worked straight through the entire program of basic steps and remedial loops with no opportunity to skip any of the remedial loops.

All group experiments were conducted by the research staff with the regular teachers observing. Students were orally urged to learn as much as they could, but all specific instructions as to learning method were printed on the program.

All students were given a criterion test of amount learned from the program. Test items were heterogeneous in form and covered the learning objectives representatively as nearly as we could judge. Appendix A shows the complete criterion test on Sets used in Experiments IV through VIII. Earlier versions were quite similar.

Effectiveness of a learning method was judged by criterion test score and time spent learning. Analyses of covariance of criterion test scores, with some standard ability test as the control variable, provided significance tests for group differences and also yielded adjusted criterion test means. A mathematical ability score was used with the Sets program and verbal ability or IQ score with the U.N. program, these choices being based on emperical correlations with the criterion tests.

As an example of how results from a study of this design should look if bypassing were the most efficient learning method, the results of our earliest bypassing study with the Sets program (in an 8th grade class) are reproduced from the previous report (Campbell,1961) and shown in Figure 1. Note that the Bypass and Long groups learned appreciably more (about 14% of the test scale from chance to perfect) than the Short group, and that the Bypass group learned as much as the Long group in substantially less time. Never since has the evidence unequivocally favored bypassing with this learning topic.

The same report describes a replication of this study in a different school and with a Medium linear method added to the design to see how much more the Bypass group would learn than a linear group taking about the same total amount of time. Learning times corresponded to those in the first study, but there were no significant differences among groups on the criterion test. Experiment I, our first experiment subsequent to that report, was a replication of the Bypass-Long-Short-Medium comparison with two classes in a third school, but with the sets lesson divided in two and given on two successive days rather than all at once. We thought that satiation midway through learning may have washed out method differences in the previous study, a hypothesis supported indirectly by disproportionately low subscores on ideas covered late in the program. Splitting the lesson, however, did not change the criterion test results (no differences among methods), and it greatly reduced learning time differences between methods. The detailed procedures, analyses and results of Experiment I, and all other studies discussed hereafter, are recorded in Part Two of this report.

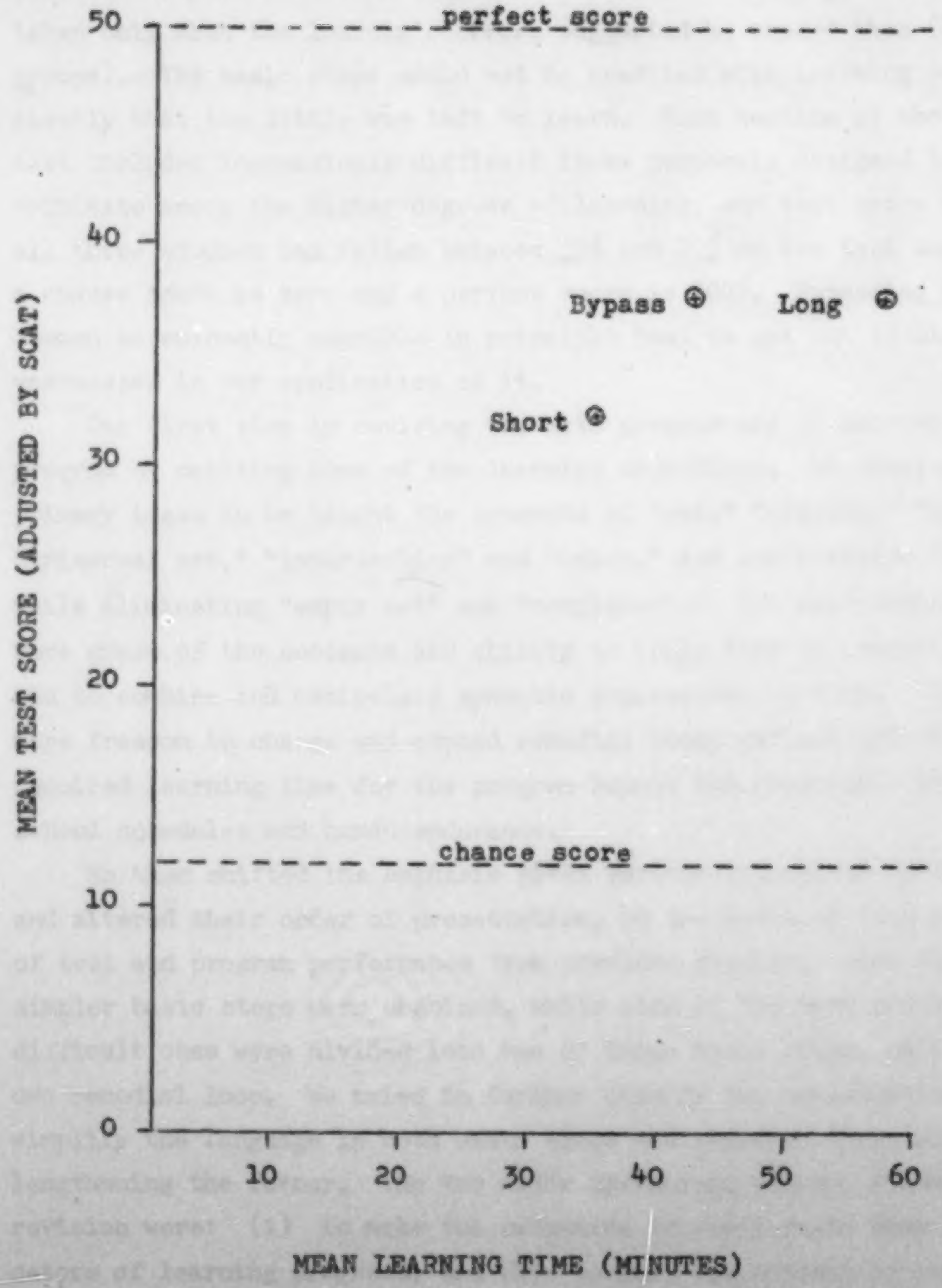


Fig. 1. Amount learned as a function of time spent on the learning program.

Our remedial loops had failed to teach more than was learned from the basic steps alone. This was true whether the loops were given to all learners as part of a full length linear program (Long groups) or were taken only when the learner's errors suggested he needed them (Bypass groups). The basic steps could not be credited with teaching so efficiently that too little was left to learn. Each section of the criterion test included increasingly difficult items purposely designed to discriminate among the higher degrees of learning, and test score means in all three studies had fallen between 30% and 70% on the test scale, where a chance score is zero and a perfect score is 100%. Bypassing still seemed so eminently sensible in principle that we set out to diagnose weaknesses in our application of it.

Our first step in revising the sets program was to shorten the basic program by omitting some of the learning objectives. We retained as primary ideas to be taught the concepts of "set," "element," "subset," "universal set," "intersection" and "union," and combinations thereof, while eliminating "empty set" and "complement." The main objectives were grasp of the concepts and ability to apply them to concrete situations and to combine and manipulate symbolic expressions of them. This gave us more freedom to change and expand remedial loops without extending the required learning time for the program beyond the reasonable limits of school schedules and human endurance.

We then shifted the emphasis given various objectives by the program, and altered their order of presentation, on the basis of item analyses of test and program performance from previous studies. Some of the simpler basic steps were combined, while some of the more complex or difficult ones were divided into two or three basic steps, each having its own remedial loop. We tried to further clarify the presentation and simplify the language in both basic steps and remedial loops without lengthening the former. The two major systematic changes attempted in the revision were: (1) to make the responses in every basic step good indicators of learning progress, and (2) to make the content of every loop more directly relevant to its basic step and more completely oriented toward a student who has failed that basic step.

One feature of the first version retained in the second was the bypassing format. The basic step took up the top $\frac{1}{4}$ to $\frac{1}{2}$ of each $8\frac{1}{2}$ x 11 page, and the remainder of the page contained the remedial loop for that step. When a student saw that his answer matched the correct one on the back he was simply directed to skip to the top of the next page where the next basic step appeared.

After preliminary tryouts of the second Sets program in Experiment II, which concerned mainly the United Nations program, we conducted Experiment III in the Bypass-Long-Short design with the second Sets program in a 9th grade class. The criterion test was given 24 hours after the program in this and all subsequent experiments. The results were strange. As a group the Bypass students scored higher on the criterion test than either the Long or Short group by about 8% (not significant), and they averaged only 2 minutes less time in learning than the Long group. We did no analysis of covariance because the correlations between Numerical Ability score on the Differential Aptitude Test (DAT) and criterion test score varied from - .68 for the Long group to + .82 for the Bypass group. This was the first study to yield markedly heterogeneous correlations suggesting that the relative effectiveness of bypassing depended on the ability level of the students. These and other ability differences are discussed in detail in the next section.

Experiment III suggested a benefit from bypassing, but not clearly, and once again we chose to place the blame upon our efforts at remedial instruction rather than upon the bypassing principle. This time we based our revision of the program on individual tape-recorded interviews with students at all ability levels, followed by administration of a criterion test. Every student understood some basic steps and every basic step was understood by someone. The main impression gained from the interviews was that if a student did not understand the short basic step, it took a great deal of questioning, discussion and illustration to get that idea across, and even this often failed. We therefore drastically revised the remedial loops so that in our third Sets program each loop was on the average six times as long as its basic step rather than one to two times as long, as was the case in the first and second versions. This required the following change in format: each step was a booklet rather than a single page. If a student correctly responded to the basic step on the first page, he skipped the rest of the booklet (the remedial loop) and began the next one.

Other changes suggested by the interviews and by related literature were:

(1) In teaching complex combinations, begin at an elementary level and take a slow part-by-part approach, only gradually combining the simpler ideas into a whole (Gagne, 1962).

(2) Avoid failures early in the remedial loop (Valentini, 1951). Application of this principle took the form of requiring no responses at all, or only easy ones, in the first two or three pages of the loop.

(3) Make a special point of discriminating and contrasting easily confused ideas and symbols, e.g. intersection (∩) and union (∪).

(4) Some implicit learning objectives are the most difficult of all, e.g., learning to ignore irrelevant sets and to use parentheses correctly. New basic steps with long remedial loops were introduced solely to teach these objectives. (The "universal set" concept was deleted to compensate for added learning time.)

(5) Series of illustrations and examples are more effective for learning concepts and principles than are abstract or formal statements during initial learning (Hendrix, 1946; Haslerud & Meyers, 1958).

After initial tryouts and minor revisions, the third Sets program was reproduced in Bypass, Long and Short forms for use in subsequent experimentation. It has also been reproduced as a separate supplement to the present report (Campbell, Terry and Shearer, 1962).

In addition to specific hypotheses to be discussed later, our final series of experiments sought to assess the overall effectiveness of bypassing with the third Sets program. Since modern mathematics is often considered as appropriate for primary school as for high school and college, and since age is the core dimension in a complex of related personality, maturation and ability variables, grade level (4th to 12th) was a major independent variable in these studies. A very practical reason for investigating a wide age range was the economic implication of being able to use a single bypass program at all grade levels in place of a more expensive series of graded texts, at least for some educational materials.

The first four experiments (IV, V, VI and VII) in this series utilized grades 4, 8 and 12, with a total of about 50 students in each grade. The final large-scale study (Experiment VIII) employed grades 4, 6, 8, 10 and 12 with about 150 students at each grade level. Bypass, Long and Short learning methods were used at each grade level except grades 10 and 12, in which the Long learning method was omitted from the design for administrative reasons. The criterion test was given one day after the program.

Experiments IV to VII suggested that bypassing worked substantial benefit at the 4th grade level (a difference of 12% on the test scale between Bypass and Short groups, $P < .05$) and that this gain diminished gradually to zero as one progressed up through the grades. Figure 2 shows the adjusted test means plotted against learning times. Only in grade 4 did bypassing appear to be more efficient than the linear methods. As might have been expected, younger students learned less by any method than older ones.

Experiment VIII with larger Ns provided more stable estimates of efficiency of the different methods at each grade level. Figure 3 shows quite clearly that bypassing was no more efficient than the linear learning methods at any grade level. Pooling all grades, the Long groups learned more than the Bypass groups by about 4½% of the test scale (from chance to perfect) and about 7½% more than the Short groups. These overall method differences, though statistically significant ($P < .001$), are disappointingly small in view of the fact that the Bypass and Long groups, averaged together, spent twice as long in learning as the Short groups. The benefit of the remedial loops is clearest in the 8th grade where the Long group scored higher than the Short group by 16%. But the Bypass group fell midway between Short and Long, both in time and test score (see Figure 3), as one might have expected a Medium linear group to do. Thus,

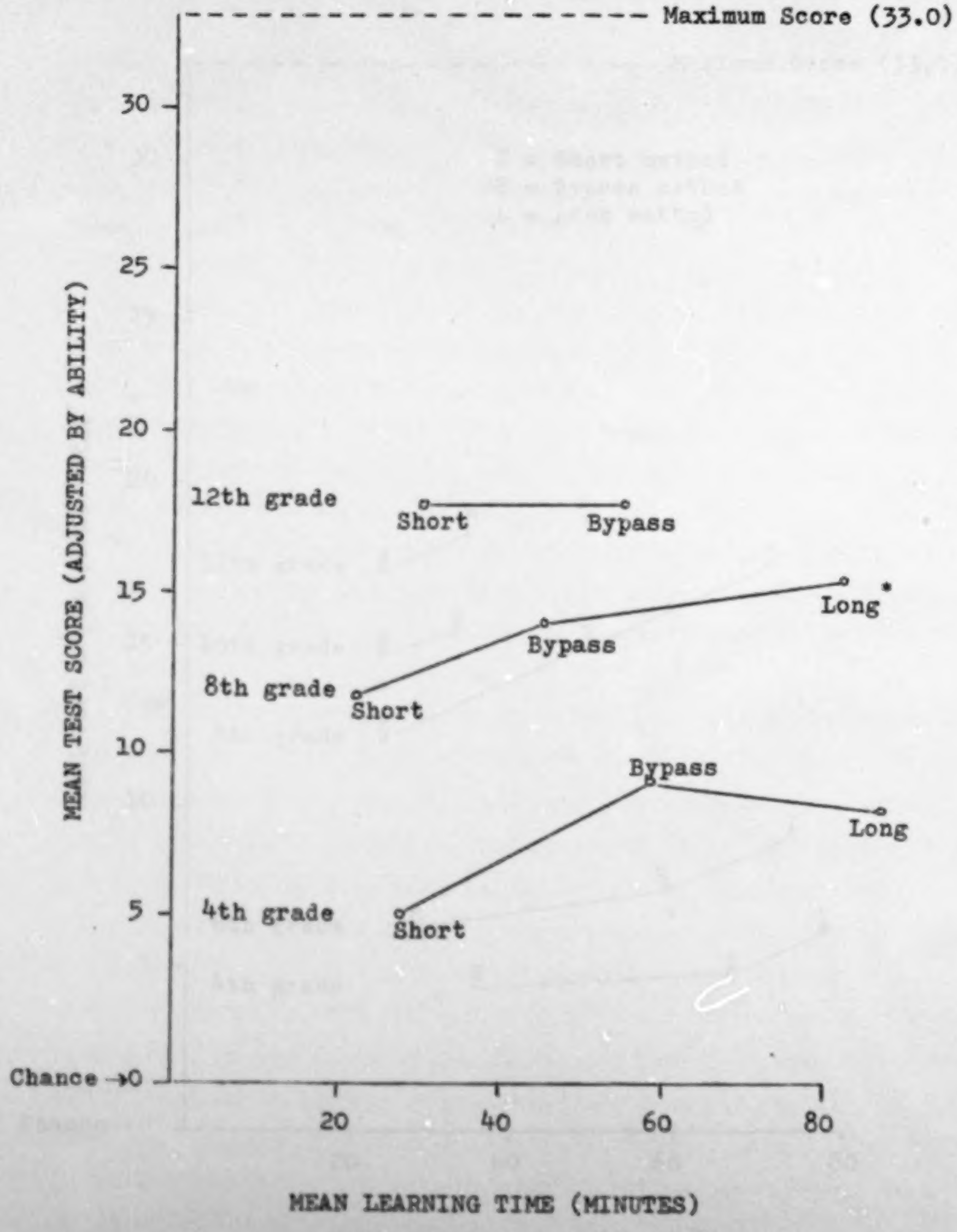


Fig. 2. Amount learned from Sets program as a function of time spent learning. Exps. IV, V, VI, and VII.

---

*This Long group is from a different school than the Short and Bypass groups.

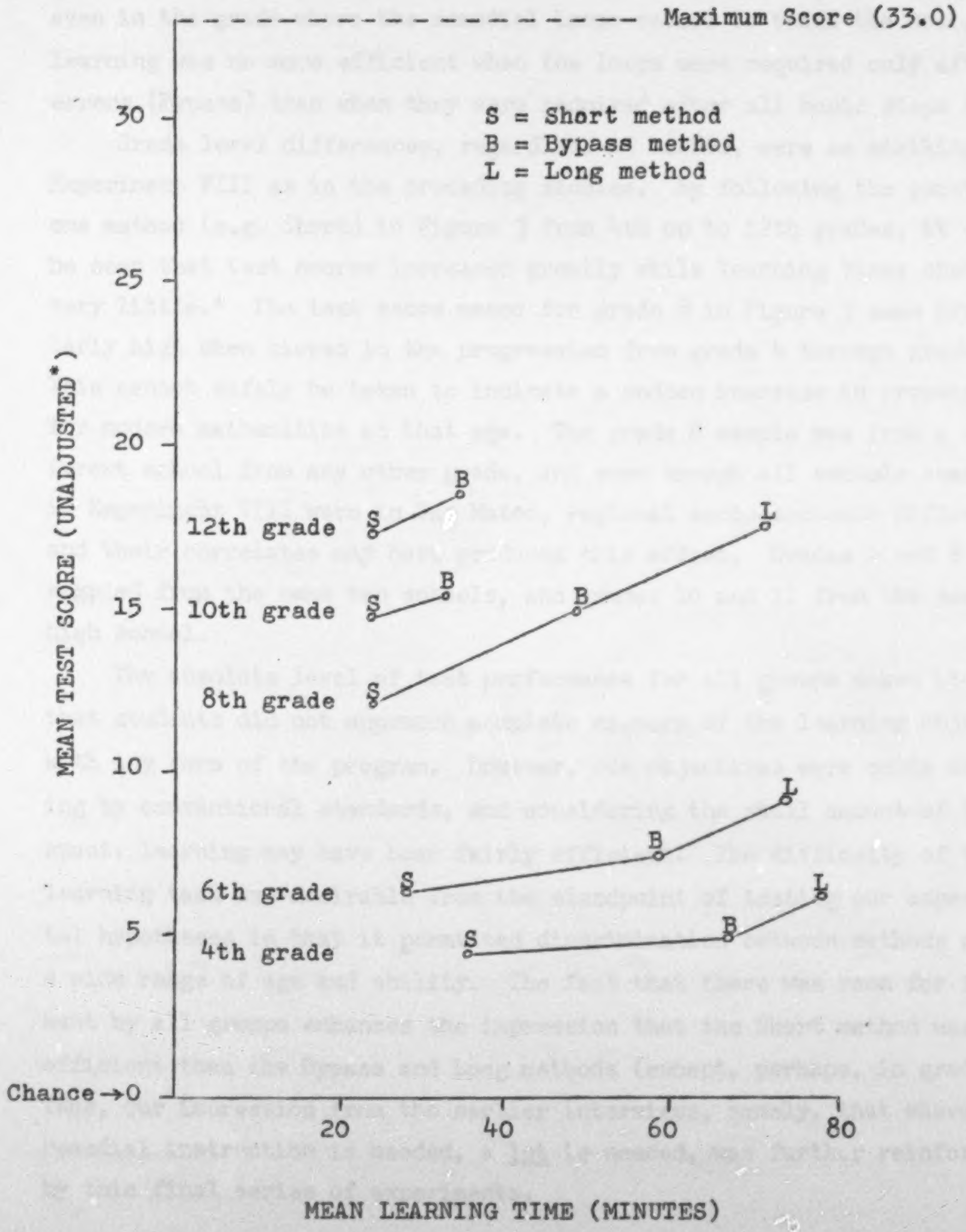


Fig. 3. Amount learned from Sets program as a function of time spent learning. Exp. VIII.

[*] Ability-adjusted means are not perceptibly different from those shown (see Part Two).

even in the grade where the remedial loops seemed to teach the most, learning was no more efficient when the loops were required only after errors (Bypass) than when they were required after all basic steps (Long).

Grade level differences, regardless of method, were as striking in Experiment VIII as in the preceding studies. By following the points for one method (e.g. Short) in Figure 3 from 4th up to 12th grades, it can be seen that test scores increased greatly while learning times changed very little.* The test score means for grade 8 in Figure 3 seem irregularly high when viewed in the progression from grade 4 through grade 12. This cannot safely be taken to indicate a sudden increase in propensity for modern mathematics at that age. The grade 8 sample was from a different school from any other grade, and even though all schools sampled in Experiment VIII were in San Mateo, regional socio-economic differences and their correlates may have produced this effect. Grades 4 and 6 were sampled from the same two schools, and grades 10 and 12 from the same high school.

The absolute level of test performance for all groups makes it clear that students did not approach complete mastery of the learning objectives with any form of the program. However, our objectives were quite demanding by conventional standards, and considering the small amount of time spent, learning may have been fairly efficient. The difficulty of the learning task was desirable from the standpoint of testing our experimental hypotheses in that it permitted discrimination between methods over a wide range of age and ability. The fact that there was room for improvement by all groups enhances the impression that the Short method was more efficient than the Bypass and Long methods (except, perhaps, in grade 8). Thus, our impression from the earlier interviews, namely, that where remedial instruction is needed, a lot is needed, was further reinforced by this final series of experiments.

---

[*] Learning times were lower for higher grades with the Bypass method partly because the older students bypassed more of the remedial loops. Analysis of Experiment IV results showed that learning time was closely proportional to number of pages read. These learning time differences are discussed further in the section on classroom effects.

Mathematical Ability of the Learner

Heterogeneous correlations between criterion test score and mathematical ability score in Experiments III, IV and V led us to suspect that the relative effectiveness of the Bypass, Long, and Short learning methods depended on the learner's ability level within his own grade. When the samples were divided at the national median ability score (at the class median in the 4th grade) and the method comparisons were made separately for high and low ability groups, an odd pattern of ability-by-method interactions emerged. Briefly, bypassing seemed to be better for high ability students in grades 4 and 9, better for low ability students in grade 8, and not much good for either ability group in grade 12 (see Experiments III to VII in Part Two for exact data).

Although these baffling trends were of questionable reliability statistically, they led us to search for similar effects in the final large-scale study (Experiment VIII). As in most previous studies, correlations between mathematical ability scores and test scores were moderately high (.40 to .70) and were homogeneous, suggesting that no ability-by-method interaction would be found. Nevertheless, each grade in the sample was divided into three approximately equal-sized ability groups on the basis of available ability scores. The California Achievement Test (CAT) Total Arithmetic score was used in the 4th and 6th grades, DAT-Numerical in the 8th, 10th and 12th grades. Although there may have been school differences, as mentioned previously, we were careful to choose our samples within a school so that they would be representative for that school population (100% samples at the 4th and 6th grade levels).

Both criterion test scores and learning times were analyzed with ability group as an independent variable, which required correcting for disproportionate Ns. The test score analyses gave no indication at all that the effectiveness of bypassing relative to the linear methods changes as a function of the student's level of mathematical ability. There was a slight tendency ($P < .05$) for high-ability students to finish the program sooner. This difference occurred mainly in the Bypass method. It makes sense that the more able students would tend to make fewer errors and thus bypass oftener and finish sooner. There were no appreciable time differences betweeen ability groups for the linear methods.

Although relative learning efficiency of the Bypass, Long, and Short methods did not depend appreciably on ability level, it is interesting to note that for all methods there were substantial differences between ability levels in amount learned. Figure 4a shows how mean test score increases with grade level and ability level for the Short method students. Very similar curves for the Bypass and Long methods are shown in Figures 4b and 4c. Note that the superiority of the high ability students over the medium and low ability groups steadily increases ($P < .001$) as grade level gets higher. A partial explanation for this is found in the answers to the question "Have you ever studied sets in school before?" Only 4% answered "yes" in the 4th and 6th grades and these were scattered across ability levels, while in the 8th, 10th and 12th grades 14% had studied sets before and 3/4 of these were in the high ability group. The students having studied sets before, as a whole, scored about 8% higher on the test than other students of the same ability group.

The Evaluation Part of Bypassing

Validity of the bypassing criterion. If bypassing is to be more than a game, a necessary requirement is that the criterion for evaluating the learner's progress be reasonably valid and accurate. Our negative findings regarding the overall effectiveness of bypassing with the Sets program add to the importance of this problem, since the evaluations in the basic steps may have been partly to blame.

It is difficult to determine the validity of individual basic step responses in the Sets program, because the ideas are highly interrelated and the test was designed to assess understanding and transfer, not rote memory of words or formulas. But since the program as a whole was meant to cover the same objectives as the test, the correlation between total errors on basic steps of the program and total test errors should be a fairly good gross index of the validity of the basic step responses as

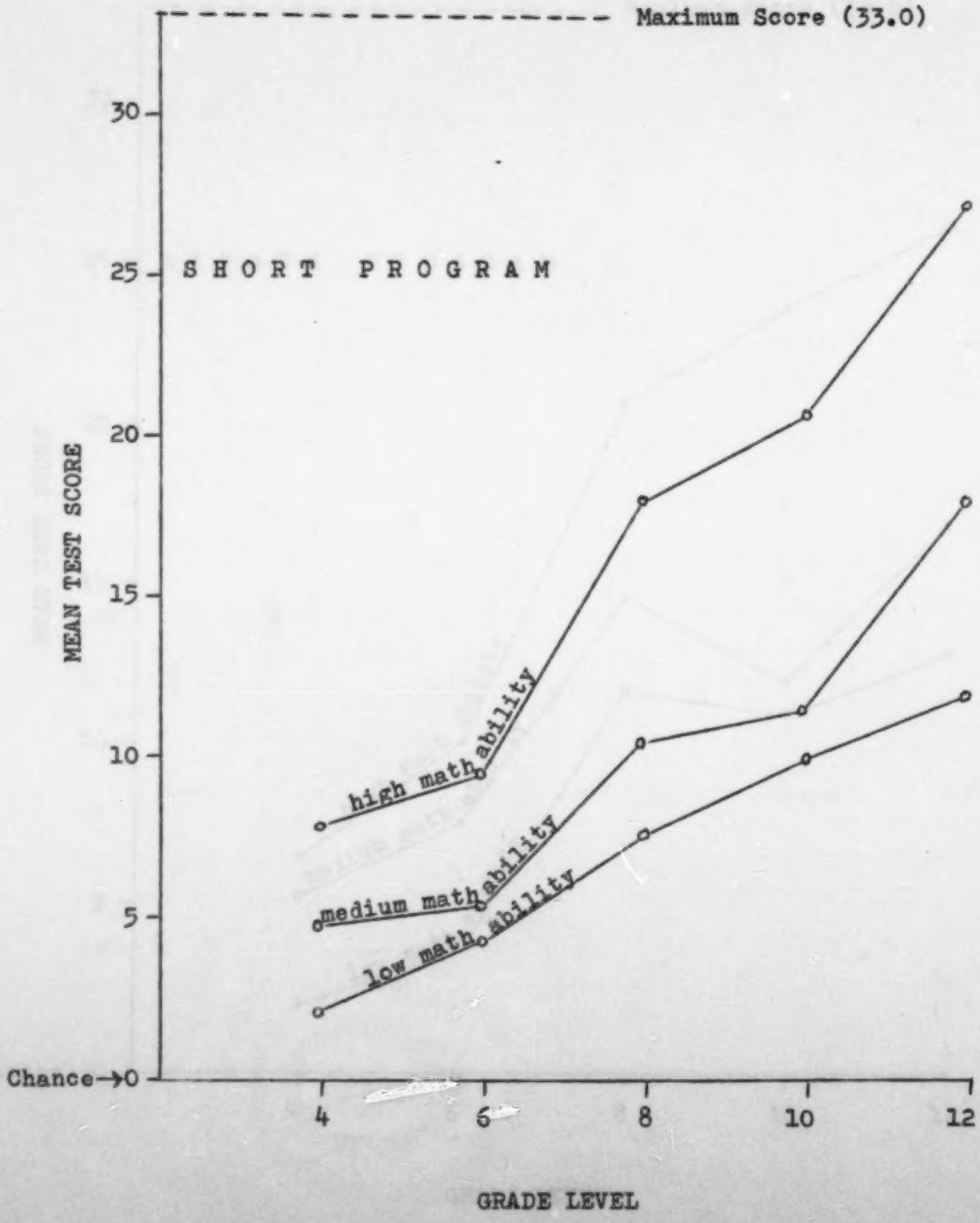


Fig. 4a. Mean score on Sets test for Short groups at three ability levels and five grade levels. Exp. VIII.

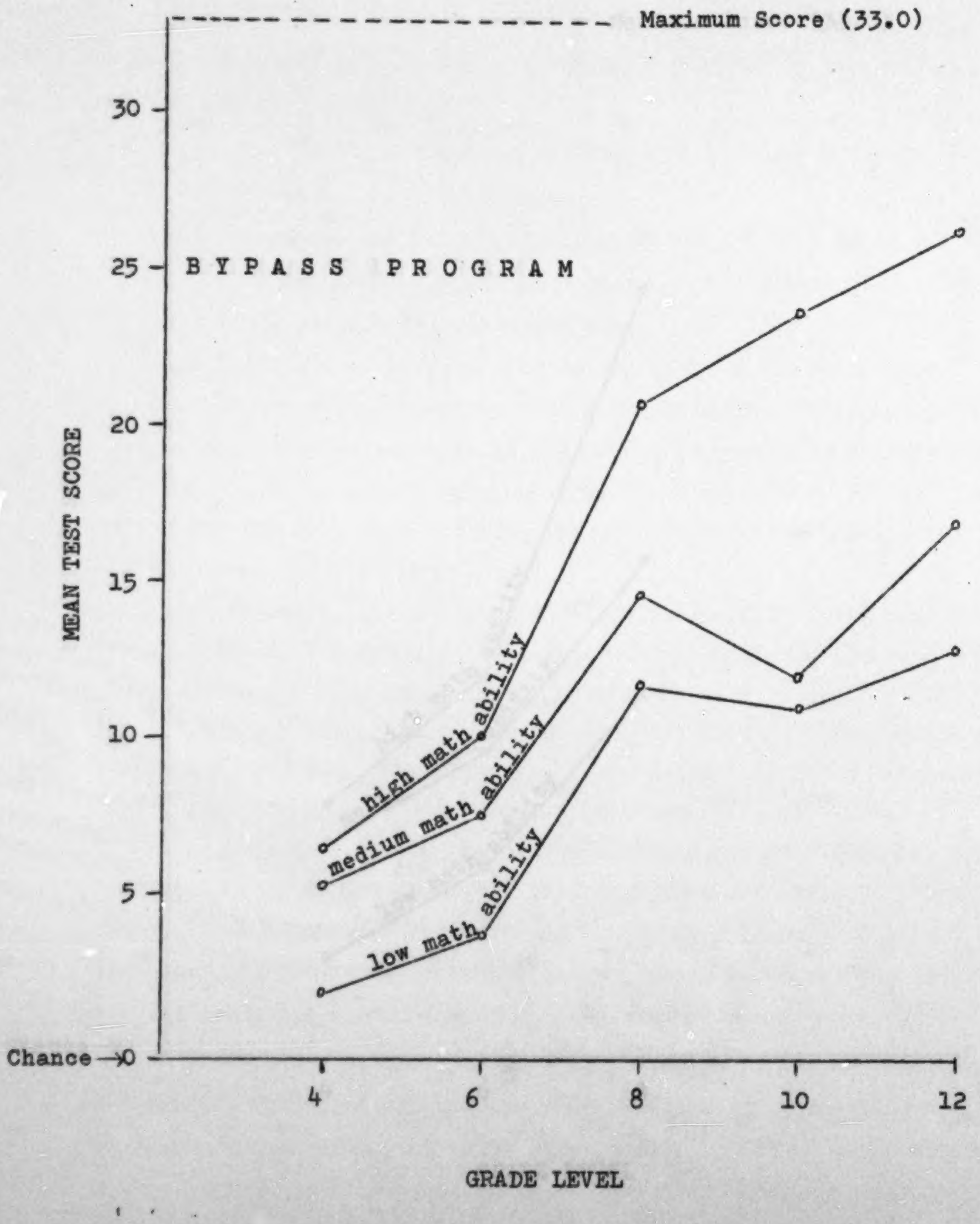


Fig. 4b. Mean score on Sets test for Bypass groups at three ability levels and five grade levels. Exp. VIII.

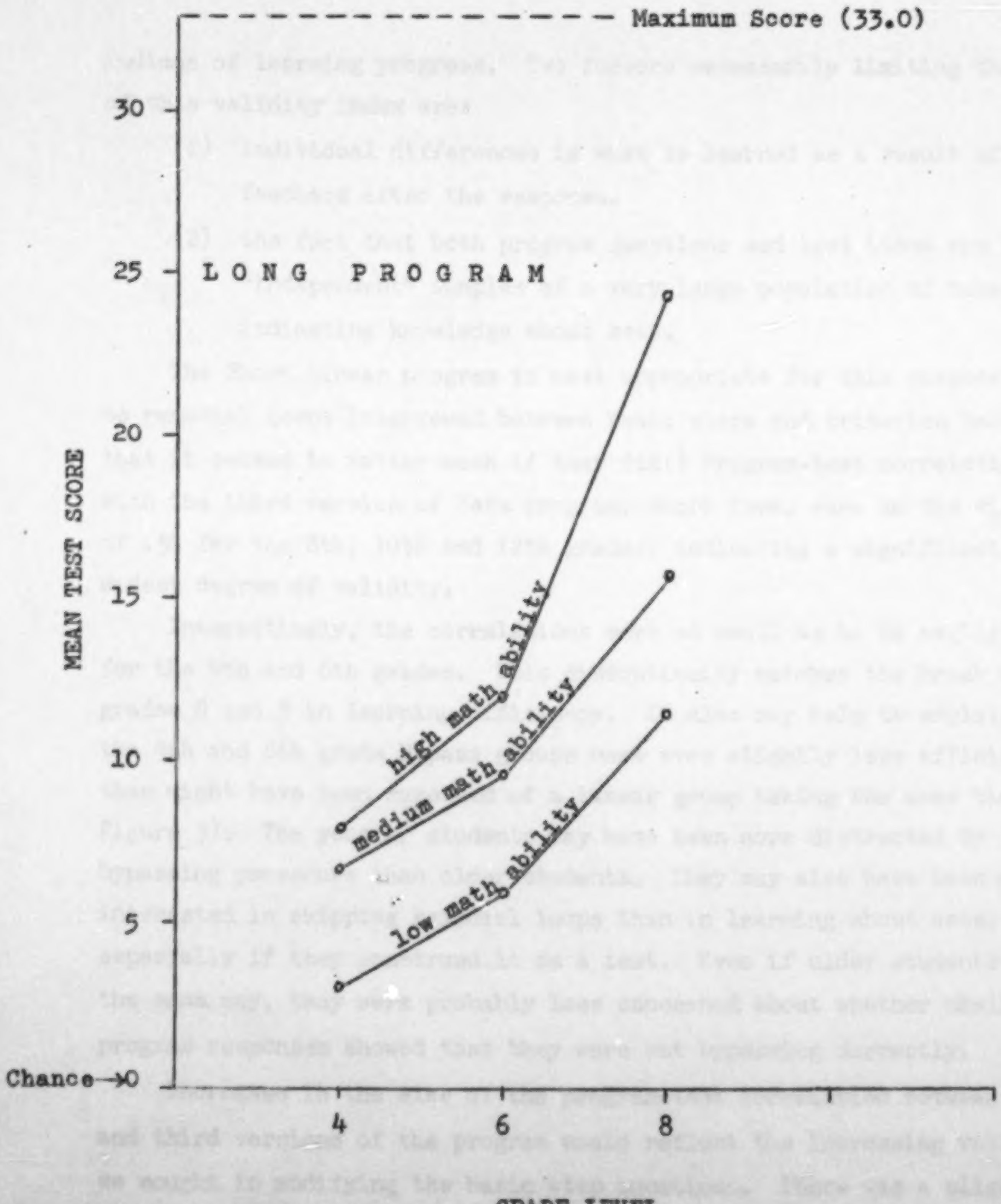


Fig. 4c. Mean score on Sets test for Long groups at three ability levels and five grade levels. Exp. VIII.

indices of learning progress. Two factors necessarily limiting the size of this validity index are:

(1) individual differences in what is learned as a result of getting feedback after the response.

(2) the fact that both program questions and test items are small "independent" samples of a very large population of behaviors indicating knowledge about sets.

The Short linear program is most appropriate for this purpose since no remedial loops intervened between basic steps and criterion test (not that it seemed to matter much if they did!) Program-test correlations with the third version of Sets program, Short form, were in the vicinity of .50 for the 8th, 10th and 12th grades, indicating a significant but modest degree of validity.

Interestingly, the correlations were so small as to be negligible for the 4th and 6th grades. This discontinuity matches the break between grades 6 and 8 in learning efficiency. It also may help to explain why the 4th and 6th grade Bypass groups were even slightly less efficient than might have been expected of a linear group taking the same time (see Figure 3). The younger students may have been more distracted by the bypassing procedure than older students. They may also have been more interested in skipping remedial loops than in learning about sets, especially if they construed it as a test. Even if older students felt the same way, they were probably less concerned about whether their program responses showed that they were not bypassing correctly.

Increases in the size of the program-test correlation between first and third versions of the program would reflect the increasing validity we sought in modifying the basic step questions. There was a slight but not significant increase in this direction. Effective remedial loops should also tend to make the program-test correlation lower for Long and Bypass groups than for Short groups. They did tend to be slightly lower, but again the difference was not appreciable.

There were fewer errors proportionally on the program than on the criterion test. This suggests another factor that may have limited the validity of the program score: Perhaps the basic steps were demanding enough to require momentary grasp of the idea, but did not require enough practice and thinking through to insure retention and transfer to new problems. Another attenuating factor was the frequency of procedural errors in which the learner bypassed after a wrong response, or failed to bypass after a right answer. About 20% of all bypassing decisions were not made according to instructions. How many of these were intentional acts of the bored or curious we do not know.

In any case, the program responses appear to have been a partially valid criterion for bypassing, which should permit bypassing to increase learning efficiency at least minimally if the topic of sets is amenable to bypassing and one has effective remedial loops. The failure of bypassing cannot be blamed solely on the bypassing criterion, it seems.

In addition to correctness of program responses, there is another obvious and economical criterion for bypassing, namely, the learner's own judgment as to how well he has learned. We compared these modes of evaluation in the 8th grade sample of Experiment IV. In addition to the group bypassing in the customary way, another group bypassed each remedial loop only if they both answered correctly and "understood the page." Since program responses were apparently easier than test items, and since "understanding" might be subjective enough to be swayed by motivation, we thought it more promising to require both understanding and response correctness than to make understanding alone sufficient for bypassing. There was no difference between the two groups in either test score or learning time. The new bypassing criterion had no apparent advantage and had the disadvantage of being more complicated, so we discontinued study of it.

Frequency of evaluation. The differences among instructional methods on this dimension are extreme. Branching programs and personal tutors typically evaluate learning after each major idea or item of information, sometimes oftener, while conventional remedial courses usually fix the learner's course of instruction for weeks or months at a time in one decision.

It might seem at first glance that one has nothing to lose by frequent evaluation, as in programs, and perhaps much to gain. But unless the evaluations are shown to be highly instructive in their own right, the time and effort expended in evaluation must be weighed against the gain in learning efficiency achieved by making instruction adaptive to the learner's step-by-step progress.

One factor in deciding how often to evaluate and branch is the structure of the learning topic, to be discussed in a later section. Another consideration is the cost of evaluation in dollars and learning time. The modes of evaluation used in the Sets program are quick and cheap, (compared to oral questioning, computerized evaluation, etc.) which makes them economical only in proportion to their validity, of course.

A third factor is the effect on learning of the evaluation itself. Whether the evaluation is educationally beneficial or neutral matters little if it is quick, but we were concerned with the possibility that a question difficult enough to be a valid index of learning, and referring only to a concise basic step, might have an adverse effect on subsequent remedial learning. This hypothesis seemed especially worth testing with the Sets program in view of the general observation that initial incomprehension of something mathematical often induces panic, indifference, hostility, or some other unreceptive state of mind, which in many people becomes a permanent attitude toward mathematics.

In order to test this hypothesis we prepared "reversed" Long programs in which each remedial loop was studied before its basic step rather than after it. We then compared groups who took this reversed Long method with groups who took the Long program in its normal order (basic step, then loop). Both methods presented the same learning materials, but the reversed Long method led the student gradually from the simple to the difficult, hopefully thereby avoiding adverse effects of immediate failure on that segment. The two methods were compared in Experiments V and VII involving 4th and 8th graders respectively. There were no appreciable differences between the two methods in test score or learning time at either grade level. Apparently the basic step did not detract from remedial learning with the Sets program.

Classroom Effects on Learning

Our experimental use of programed instruction in the schools was not intended to be a model of how programs should be used in practice. We did not approach the problems of integrating programs with other methods, nor how best to utilize time made available by bypassing. However, we gave some attention to the problems of timing and social distraction as possible extraneous variables in our method comparisons.

Experiment I was premised on the possibility that satiation might have artifically obscured the advantage of bypassing in our early studies. As noted before, we found no evidence of such an effect. The opposite side of the time problem is that limiting the time might have had a differential effect on the learning methods since the Bypass method required roughly twice as long, and the Long method three times as long, as the Short method. The percent of students in grades 4, 6 and 8 (Experiment VIII) who did not completely finish the program in the 100 minutes available was small (8%). In the 10th and 12th grades, where the time limit was 50 minutes and no Long programs were given, only 9% did not finish. These students were included in all analyses. We have two empirical indications that the time limit had no appreciable effect on criterion test scores. First in Experiment VIII, and previous studies as well, correlations between time spent learning and test score averaged zero or near-zero for all learning methods. A reasonable interpretation of this is that individual differences in motivation would make the correlation positive and ability differences would make it negative, but the two effects cancel each other to make the correlation zero.

Comparison of the 12th grade findings in Experiments VI and VIII provides a second indication that time limits did not differentially reduce test scores for the methods requiring more time. Mean test scores were almost identical for the two Bypass groups even though the Experiment VI sample had more time available and took more time (55 minutes vs. 35 minutes), while test score means were equivalent for the two Short groups, and their learning times differed by only six minutes.

In light of this evidence, time pressure is not a good explanation for the finding that high school students taking the Bypass program apparently gained little from the remedial loops, while younger students gained more. A likelier reason is the one mentioned earlier: high school students bypassed more of the loops than younger students did.

All eight experiments reported were conducted in classrooms or similar group settings. Since time of finishing the program varied greatly among individuals, and systematically among methods, we wondered whether the activities of students who had finished the program would distract those who had not. We were especially concerned about the high school sample in Experiment VIII, because they were assembled in groups of 40 to 80 for the experiment, and we had previously observed that general commotion and bustle increases with the size of the group. In order to determine whether gradually increasing background noise was diluting the amount learned, by the Bypass method in particular, one of the three 12th grade rooms (50 students) was kept in almost complete quiet throughout the learning period (perfect silence, like a perfect vacuum, is not practically attainable). The resulting test score means showed no sign at all of greater learning in the quiet room, either for the group as a whole, or for the Bypass group relative to the Short group.

One other comparison bearing on social distraction effects is in order. As students completed the Short Sets program in Experiments IV, V and VI, they were given other learning materials on a different topic which occupied most of the remaining time. No such diversion was available for the early finishers in Experiment VIII. If social distraction detracted from learning, the superiority of the Bypass group over the Short group should be greater in Experiments IV, V and VI than in Experiment VIII. One can see by comparing corresponding grades in Figures 2 and 3 that the Bypass groups as a whole surpassed the Short groups by no more in the early experiments than in Experiment VIII.

Apparently social distraction was not a notable factor in determining amount learned.

Research with the United Nations Program

The same pragmatic orientation guided construction and revision of both the Sets program and the United Nations program. The main differences were in format and step size. The United Nations represented a more verbal and discursive learning topic than set theory, and it seemed appropriate to use at least paragraph-size steps in order to avoid loss of continuity. This created a problem in devising a bypassing version of the program in that each such unit of discourse covered at least two or three main points of information. Our answer to this in the first version was to ask from two to four multiple-choice questions in succession on a page following the basic step page. The student was instructed to skip to the next basic step only if he made no wrong answers to these questions. We used a chemical method of giving immediate feedback. A red spot appeared on the paper if the student chose the right answer. (He was instructed to keep choosing until he found it.)

As in the early versions of the Sets program, the remedial loops in the U.N. program were about the same length as the basic steps or slightly longer. The ideas were cast in a different organization in the loop (e.g. an outline or diagram instead of continuous discourse) if clarity of presentation did not appear to suffer thereby.

Effectiveness of Bypassing and Remedial Loops

Our initial experiment with the U.N. program (Exp. II) was conducted in a 5th grade class using only linear methods. We anticipated that grade level would be a variable in future experiments and wanted to determine whether our efforts to simplify the language had been successful enough to make the program appropriate for that age group. Since the evaluative questions and the loop each required about the same amount of space and study time as the basic step, we compared the following three linear forms of the program:

(1) The basic steps only with no questions.

(2) The basic steps and question pages but no remedial loops.

(3) The full-length form: basic steps, questions and loops.

There were seven short chapters to the U.N. program, and a criterion test on each was given immediately after the student completed the chapter. The main learning objective of this first version was immediate grasp of fairly specific knowledge about the U.N., and the test items were mainly questions asking for answers of word or phrase length.

Analysis of covariance of criterion test scores with CAT-Total Reading score as the control variable revealed only marginally significant ($P < .10$) differences between methods, and the pattern of means was very odd. The shortest and longest versions did equally well, but the middle-length program with only the loops missing produced lower mean test scores by about 10% of the test scale. Learning time means corresponded roughly to program length, the full-length version taking 20 minutes longer than the shortest and the middle-length version 10 minutes longer. It is hard to imagine why answering the evaluative questions would detract from the amount learned from a United Nations program, unless the students became so interested in finding red spots that they forgot about the learning topic. In any case, the combination of questions and loops given in linear form apparently added nothing in terms of amount learned. Since the mean test score of the basic-step-only group (about 30% above chance) left plenty of room for improvement, this suggests that the topic may have been inappropriate for this grade level because 5th graders lack interest in the complex social issues and institutions involved. The more difficult key words, such as "administer," "recommendation," and "trusteeship," probably only acquire meaning with years of exposure, normally. To teach these by illustration in concrete personal terms seemed too large an undertaking for our experimental purposes, so we restricted further experimentation with the U.N. program to higher grade levels (8th to 12th).

After minor revisions of the program, we conducted an experiment of the Bypass-Long-Short design with the U.N. program in a 9th grade class (Exp. III). The Short program included the basic step and question pages but no loops. Again, a criterion subtest was given after each chapter.

Using IQ (Terman-McNemar) as the control variable, analysis of covariance of test scores yielded no significant differences, the Bypass group mean only slightly exceeding the other two. The class as a whole scored considerably higher (about 60% above chance) than the 5th grade class. Learning time means were in the expected order, increasing from Short to Bypass to Long. It appeared that even at the 9th grade level little or nothing was gained by reading the remedial loops, in bypassing format or otherwise. This experiment yielded similar results with the second Sets program, as discussed earlier, and we decided that both programs (U.N. and Sets) contained too little remedial instruction in their loops.

The first step in revising the U.N. program was to change the learning objectives both in quantity and kind. A great deal of the specific factual information presented in the first version was omitted and major ideas were given more emphasis. Our criterion test was changed correspondingly to emphasize more general knowledge. The remedial loops were then rewritten and greatly expanded so that they were several times as long as their basic steps and contained additional questions and answers on the same ideas.

Another major change in the program was the arrangement of the basic questions and loop material. Instead of asking three or four questions in succession just after the basic step, each question covering a point in the basic step was followed immediately by the part of the remedial loop pertaining to that particular question. We thus retained the same "readable" size of basic step while enabling bypassing to be based on single ideas rather than aggregates of information. The chemical feedback feature was also dropped. The correct answer to each question appeared at the top of the following page.

Experiments VI and VIII included our final evaluations of bypassing with the revised U.N. program. They were conducted with 8th and 12th grade students (about 35 in each grade), and a somewhat different design was used. The Bypass groups skipped remedial loop segments according to the correctness of their answers, as usual. The Short groups studied the basic steps but were not given the evaluative questions. This treatment differed from the Short Sets program where the evaluative question was part of the basic step, but in view of the Experiment II findings, the

U.N. Short group probably would have learned no more if the questions had been included. Because of the time required to complete the full-length form without bypassing any loops, there were no Long groups in these studies. Also, considering the total lack of benefit ensuing from the earlier U.N. loops, we were most concerned with whether the Bypass groups would surpass the Short in amount learned.

The whole criterion test was given as a single unit immediately upon completion of the learning program. Since we had shifted the program and test objectives toward general knowledge in the revised version, we thought it worthwhile to include control groups getting only the criterion test, in order to establish a baseline of prior knowledge.

The mean test scores (adjusted by DAT-Verbal Reasoning) for the Bypass, Short and Control groups are plotted against their mean learning times in Figures 5 for 8th and 12th grades separately. Both learning methods clearly surpassed their no-learning control conditions. The Bypass groups apparently learned more than the Short groups in both the 8th ($P<.10$) and the 12th ($P<.05$) grades, but the difference was greater in the 12th grade. The Bypass-Short difference did not depend appreciably on verbal ability in either grade.

The bypassable remedial loops of the U.N. program seemed to have caused a tangible increase in amount learned. Whether the gain is worth the extra half-hour of learning time is a moot point in the absence of learning norms for this subject matter. We also lack evidence as to whether bypassing was more efficient than taking all remedial loops in linear fashion. Judging by the generally consistent relation between learning time and amount of material covered, bypassing would save time. But the extra time spent in the linear method might well add proportionally to the amount learned, as it did with the Sets program in Experiment VIII. To sum up, our preliminary findings with the U.N. program give us tentative hope that bypassing might increase learning efficiency with this topic at the high school level, but it should not be taken for granted considering our findings with the Sets program.

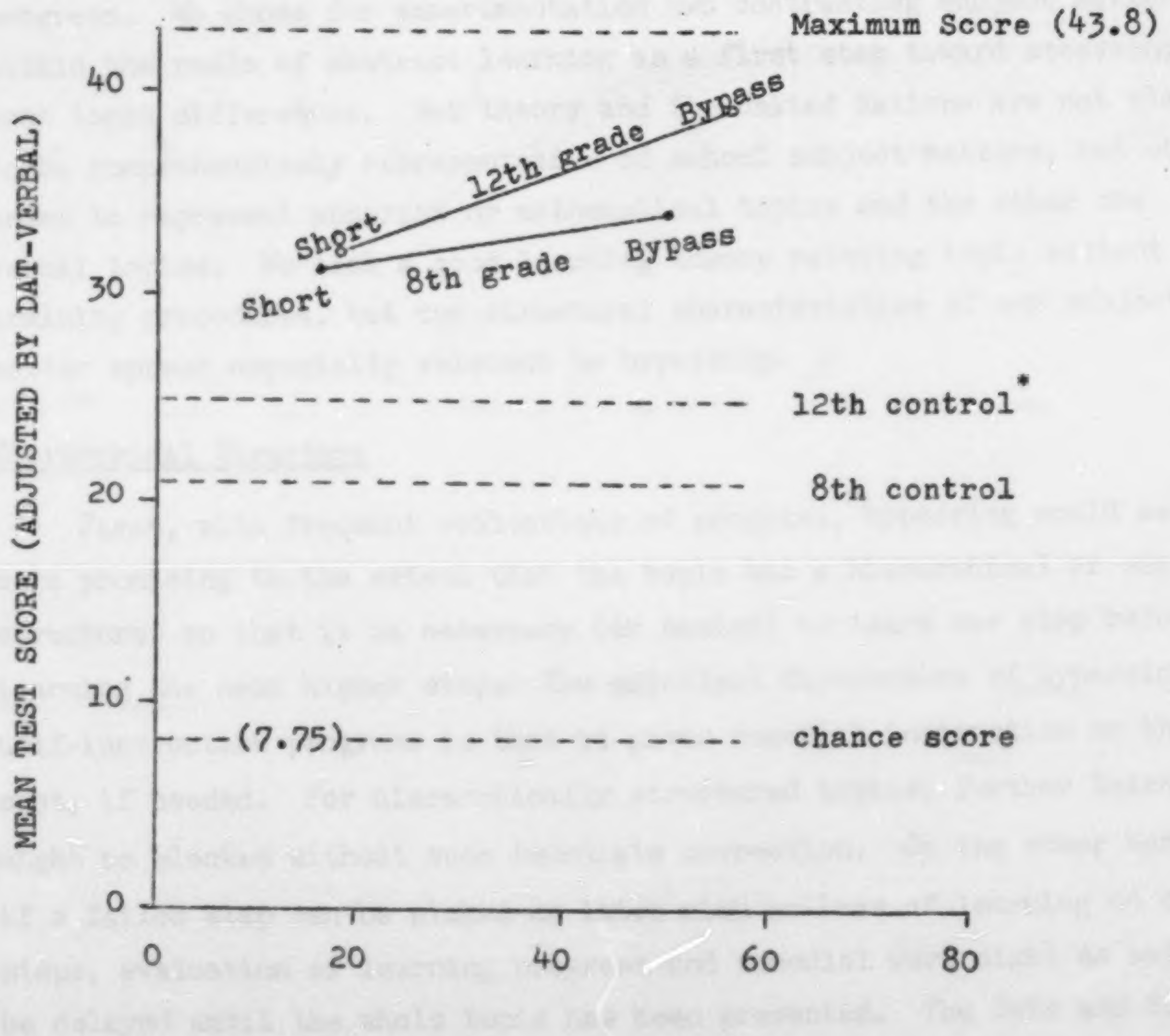


Fig. 5. Amount learned from U. N. program as a function of time spent learning. Exps. VI and VIII.

---

*The 12th grade control group was from a different school from the Short and Bypass groups; but the two schools involved are the same two in which 12th graders scored equivalently on the Sets test.

Topic Structure and Learning Objectives

As noted earlier, the nature of the topic and the learning goals may set limits on what can be gained by adapting instruction to the learner's progress. We chose for experimentation two contrasting subject matters within the realm of abstract learning as a first step toward assessing such topic differences. Set theory and the United Nations are not claimed to be comprehensively representative of school subject matters, but one seems to represent abstract or mathematical topics and the other one verbal topics. We lack a good learning theory relating topic content to training procedures, but two structural characteristics of any subject matter appear especially relevant to bypassing.

Hierarchical Structure

First, with frequent evaluations of progress, bypassing would seem more promising to the extent that the topic has a hierarchical or vertical structure, so that it is necessary (or easier) to learn one step before learning the next higher step. The principal distinction of bypassing in self-instruction programs is that it gives remedial instruction on the spot, if needed. For hierarchically structured topics, further learning might be blocked without such immediate correction. On the other hand if a failed step can be picked up later with no loss of learning on other steps, evaluation of learning progress and remedial work might as well be delayed until the whole topic has been presented. The Sets and U.N. topics both appeared to contain hierarchies of ideas, but no quantification of this property was attempted.

Difficulty Interdependence

A second subject matter characteristic which we did quantify is the extent to which different learners stumble on different ideas. If the rank order of difficulty of the steps to be learned is the same for everyone, even though some learners have more trouble overall than others, then one good early evaluation (or even a pretest) should suffice for predicting how much instruction a person needs throughout the topic. On the other hand, if persons vary greatly as to which steps they find difficult to learn, then it would seem desirable to evaluate the learner's mastery of each part separately and give remedial instruction accordingly, that is, to use the bypassing technique.

We developed a statistical index similar to Guttman's coefficient of reproducibility (Green, 1954) which quantitively expresses this independence of difficulty of various parts of a learning topic for a given sample of learners. The index is described in detail in Appendix B. Briefly, given N students and K items of a test or program, the index reflects the extent to which one can predict the NK responses by knowing only the rank order of difficulty of the items and each person's total score. A value of 1.00 indicates that the rank order of item difficulties was the same for all learners, suggesting that bypassing would be no better than, say, having every learner review or repeat the whole lesson (with difficult steps properly emphasized) until he meets a criterion of mastery. A value of zero indicates no relation between the patterns of errors of different students, and bypassing should have higher potential for such a topic.

This index (D) was obtained for the two experimental topics using item analyses of program responses from earlier studies with Sets and U.N. programs. Average value of D for the U.N. program was .25, for the Sets program .65. On the basis of independence of step difficulties alone, it appears that the U.N. topic may be better suited to bypassing than is set theory. Our general findings on bypassing with the U.N. and Sets programs suggested the same possibility, but not clearly enough to be called corroboration.

Whatever the topic, it may be that the efficiency of bypassing depends on whether the learner is expected to memorize, to understand, to transfer knowledge to new situations, and/or many others. One especially intriguing question is "Would bypassing be more effective in later stages of learning than in initial attempts to grasp and relate ideas?" The bypassing studies of the present project primarily involved initial learning, and with the Sets program we concluded that even greater amounts of remedial instruction were needed. Perhaps bypassing could be efficient with remedial loops of reasonable length if students were reviewing or relearning these concepts rather than forming them.

Some may conclude that if educators and programers must search so hard for topics and learning objectives which are amenable to efficient bypassing programs, the technique is not worth the trouble. If one should find a type of learning task favorable to bypassing, however, the simplicity and economy of bypassing would justify its use on even a small scale. At present the most likely plan for vindicating bypassing, if it can be done, seems to be to choose a hierarchically structured learning task, but one in which learners differ more as to which parts they find difficult than was the case with set theory. One should apparently be prepared to give remedial loops of instruction which may take 20 or 30 times as long as the basic steps, or longer. If remedial instruction does demand such effort, it may prove best to begin by identifying what relevant fundamentals the learner does know and what misconceptions he has.

Whatever the outcome of such research the general problem of adaptive instruction cannot be ignored. There are certain core subject matters in which we want all students to learn the fundamentals. Branching and bypassing in self-instruction programs may not prove to be effective for any given one, but the problem of how those who fail initially can learn to a minimum criterion of mastery will remain.

PART TWO

Record of Variables, Procedures, Analyses and Results, in Eight Bypassing Experiments

Experiment I (September 1961)

Independent Variable

Bypassing vs. linear forms of first Sets program, the program being given in 2 segments to avoid satiation. The 4 learning methods:

Bypass: S takes all basic steps, but only those remedial loops where his response on the preceding basic step was incorrect.
Long (linear): S takes all basic steps and all remedial loops.
Short (linear): S takes all basic steps but no remedial loops.
Medium (linear): S takes all basic steps and certain remedial loops predetermined by E (experimenter); specifically, S takes those loops following the basic steps most difficult on the average for Ss in previous studies.

Dependent Variables

(1) Score on a mixed-type 50-item criterion test shown in the appendix of a previous report (Campbell, 1961) and similar to the one shown in Appendix A of this report.
(2) Numerical score on the Differential Aptitude Test (DAT).
(3) Program score -- number of correct responses to 39 basic steps.
(4) Learning time (minutes spent on the program).
(5) Testing time.

Subjects: Two 9th grade algebra classes (N=54)

Procedure

Ss were assigned quasi-randomly to one of the 4 methods by passing out the different program forms alternately. The instruction pages for the Bypass and linear forms of the program were quite similar to those shown for Exp. III (pp.41-42). Approximately the first 2/3 of the program (all methods) was given at the first meeting. The other third of the program was given one day later and was followed immediately (as each S finished) by the criterion test. At the start of the second day all Ss were given a one-page review of the previous day's lesson.

Analyses and Results

Analyses of variance and covariance (DAT-Num.) of criterion test scores yielded no significant differences. Learning times differed significantly (P<.001) but less than in other studies reported. Testing times did not differ significantly.

Table of Means and Estimated Standard Deviations

Method	DAT-Num	Criterion test score (chance = 12.5, max. = 50) Unadjusted	DAT-adjusted	Learning Time (Min.)	Testing Time (Min.)
Bypass	65.7	25.6	26.3	50.7	20.3
Long	70.4	29.0	28.9	65.7	18.3
Medium	76.9	28.8	27.8	54.9	20.0
Short	68.0	27.9	28.2	49.4	21.6
Est. S. D. (within methods)	18.0	6.38	4.6	8.82	3.89

Average within-method correlations (Pearson r)	Sets Test	Program Score	Learning Time
DAT-Num	.55		
Program Score	.37		
Learning Time	.03	-.16	
Testing Time	-.09		-.11

Experiment II (November 1961)

Independent Variable

Type (length) of linear United Nations program:
- (1) Basic steps only (19 paragraph-size steps)
- (2) Basic steps and evaluative questions, one page of 2 to 4 multiple-choice questions following each basic step page.
- (3) Basic steps, evaluative questions, and remedial loops, one loop page following each question page.

Dependent Variables

- (1) Score on a criterion test having 43 completion items and 10 multiple-choice items.
- (2) Ability scores - California Achievement Test (CAT) - Total arithmetic
  - CAT - Total reading
  - California Test of Mental Maturity (CTMM) - IQ
- (3) Total time taken on program and test

Subjects: One 5th grade class (N = 27)

Procedure

Ss were assigned "randomly" to learning methods by alternation, as in Experiment I. The program instruction page for the two longer methods having evaluative question pages is shown following the results section below. The basic-step-only group was merely instructed to read each page carefully, and not to turn back to earlier pages. The program was divided into seven short chapters, each in a separate booklet. The subtest on a given chapter was attached to the front of the subsequent chapter, so that S was tested over each chapter just after he finished it. After the first 45-minute period of work, all Ss took a 15-minute recess, then continued on their programs. Ss who finished early were given an incomplete tryout version of the second Sets program.

Analyses and Results

Analysis of covariance (CAT-Reading) of test scores yielded a marginal ($P < .10$) difference between methods. Learning times differed significantly ($P < .001$).

Table of Means and Estimated Standard Deviations

Method	CAT-Reading	U.N. criterion test (chance = 3.2, max. = 50) Unadjusted	CAT-adjusted	Total time (Min.)
Basic	7.86	16.2	16.4	42.7
Basic + questions	8.38	14.8	10.9	52.3
Basic + questions + loop	7.46	12.9	16.2	61.7
Est. S. D. (within methods)	.921	8.61	5.18	7.21

Correlations	Total Time	CAT-Reading	CAT-Arithmetic	CTMM IQ
U.N. test	.12	.82	.53	.47
Total time		- .24		

INSTRUCTIONS

These booklets are to help you learn about the United Nations. Each booklet is one chapter. In each chapter are several "review" questions. These questions are to help you learn. After reading a question press the felt tip of your pen on the letter in front of the answer you think is right. If you choose the right answer, your pen will make a red spot on the paper. If your choice is wrong, no red spot will appear. In that case, think about the question again and make another guess.

NEVER LEAVE A QUESTION UNTIL YOU FIND THE RIGHT ANSWER.

Now, remove the cap from the pen and try it on the sample question below.

Sample Review Question: Which animal has four legs and barks? (Press the felt tip on the letter in front of the right answer.)

a. chicken

b. dog

c. cat

d. snake

Remember to read each page carefully. When you finish a page, go on to the next page and please do not turn back. When you finish a booklet, raise your hand.

IF YOU HAVE A QUESTION ABOUT THESE INSTRUCTIONS, RAISE YOUR HAND, IF NOT, TURN THE PAGE AND BEGIN THE LESSON.

Experiment III (December 1961)

Independent Variable

Bypass vs. Long and Short linear learning methods, using the second Sets program and the first United Nations program (see Exp. I for definitions of Bypass, Long and Short).

Dependent Variables

(1) Ability scores:
DAT - Verbal Reasoning
DAT - Numerical Ability
DAT - Abstract Reasoning
Terman-McNemar - IQ
(2) Sets program variables:
Criterion test - 32 items (similar to the one shown in Appendix A)
Program score - number of correct responses to the 18 basic steps.
Learning time - minutes spent on the program.
(3) U.N. program variables:
Criterion test - score on the 23-item part (about half) of the program finished by all Ss.

Subjects: One 9th grade English class (N = 22)

Procedure

On the first day Ss were assigned "randomly" (by alternation as in Exp. I) to methods for the Sets program. When an S had finished the Sets program he began working on the U.N. program, in which subtests followed chapters (see Exp. II). For time economy, Ss having Long Sets programs were given Short U.N. programs and Ss having Short Sets programs were given Long U.N. programs. Ss having Bypass Sets programs also took Bypass U.N. programs. On the second day (one 50-minute period each day) all Ss began by taking the Sets criterion test. Then each S continued on the U.N. program and subtests (the one day break came between chapters for every S so that criterion tests would not be delayed).

The instruction pages for Bypass and linear forms of the Sets program are shown following the results section below. In this second version of the Sets program, the bypassing instruction was repeated after every basic step.

The instruction page for the linear forms of the U.N. program was the same as that shown on the previous page for Exp. II. The U.N. Bypass group had approximately the same instruction page, but with the following paragraph added:

"After each page of review questions there is a yellow page of extra review. If you make any wrong guesses on a page of review questions, read carefully the yellow page which comes after that. If you make no wrong guesses with your felt-tip pen, skip the yellow page entirely."

Analyses and Results

Sets Program and Test

Analysis of variance of all test scores yielded no significant differences. However, when only Ss having scores above the 50th percentile on DAT numerical were included, method differences were significant (P< .01). The latter analysis was suggested by heterogeneous correlations between ability scores and criterion test scores. Methods differed in mean learning time (P< .005).

Table of Means and Estimated Standard Deviations

Method	DAT-Num.	Sets test (chance = 8.8, max. = 32) All Ss	High DAT Ability	Low DAT Ability	Learning Time (Min.)	Program Score (Max.=18)
Long	58.0	13.6	10.7	18.0	30.1	13.1
Bypass	59.9	16.0	17.5	14.0	27.9	12.7
Short	63.4	12.7	13.5	11.7	16.0	12.6
Est. S. D. (within methods)	28.4	3.43	2.14	3.48	6.52	

Correlations (all Ss combined)	DAT Num.	T-M IQ
DAT-Num.		.68
DAT-Verb.	.60	.78

CORRELATIONS: Computed for each method separately because of great disparities between methods. Note that entries are based on very small Ns (about 7 per entry).		Learning Time	Sets Test	DAT Num.	DAT Verbal	DAT Abst.	T-M IQ
Sets Test:	Long			-.68	-.50	-.78	-.22
	Bypass			.82	.62	.38	.46
	Short			.74	.72	.84	.46
Learning Time:	Long		.27	-.32			
	Bypass		.54	.58			
	Short		.00	-.23			
Program Score:	Long	.71	.14				
	Bypass	.38	.61				
	Short	-.42	.56				

U.N. Program and Test

Analyses of covariance of U.N. test scores with IQ as the control variable (and again with DAT-Num.) yielded no significant differences. Learning times differed significantly ($P < .005$).

Table of Means and Estimated Standard Deviations

Method	T-M IQ	U.N. Criterion part-test (chance = 2, max. = 23) Unadjusted	IQ-adjusted	Learning Time (Min.)
Long	110	13.5	14.1	48.3
Bypass	117	16.1	15.1	38.0
Short	111	14.3	14.7	31.2
Est. S. D. (within methods)	16.6	4.70	2.74	7.15

Average within-method Correlations	Total Time	U.N. Test	DAT-Num.	DAT-Verbal	T-M IQ
U.N. Test	-.31		.71	.74	.85
Total Time		-.31			-.64
Program Score	-.25	.62			

Form B (BYPASS)

Exp. III

Name ____________________

INSTRUCTIONS

This booklet is a first lesson on reasoning about sets. Every page has review exercises mixed in with the reading matter. Work these with your pencil as you come to them. Most of them are blanks in which you write in a word or signs. Or you may be asked to draw a line.

After doing any one of these things, look immediately on the back of that page to see if you answered correctly. (The right answer will be on the back of the page at the same level as the question or picture you have just completed.) Take your pencil and work the example below to be sure you understand this.

Example

Suppose that "men from Mars" all have two heads and only one foot. You come into a room full of these creatures and decide it is easier to count feet than heads. There are 12 feet in the room. From this you know that there are ________________ men from Mars in the room.

(fill in the number then look at the right answer on the back before reading on)

Whenever you work such an exercise in the lesson, be sure to check the answer on the back immediately afterward, and IF YOUR ANSWER WAS RIGHT GO DIRECTLY TO THE NEXT PAGE WITHOUT FINISHING THE PAGE YOU WERE ON. If your answer was not right, continue working on the same page. The reason for your doing this is that each page covers only one idea and goes over it 3 or 4 times. We want you to spend the most time on the ideas that give you the most trouble. So any time you check the back of a page and see that your answer was correct, skip directly to the next page.

Don't worry about mistakes; they may even help you learn. If you are not sure of an answer, guess or put a "?" in the blank. Work at your own speed, but learn as much as you can in the time you spend. You may write and draw anywhere on the booklet if it helps you to study. Tomorrow we will give you a quiz to see how much you learned from this lesson.

IF YOU HAVE A QUESTION ABOUT THESE INSTRUCTIONS, RAISE YOUR HAND. IF NOT, TURN THE PAGE AND BEGIN THE LESSON.

Form L (LINEAR)

Exp. III

Name ________________________

INSTRUCTIONS

This booklet is a first lesson on reasoning about sets. Every page has review exercises mixed in with the reading matter. Work these with your pencil as you come to them. Most of them are blanks in which you write in a word or signs. Or you may be asked to draw a line.

After doing any one of these things, look immediately on the back of that page to see if you answered correctly. (The right answer will be on the back of the page at the same level as the question or picture you have just completed) Take your pencil and work the example below to be sure you understand this.

Example

Suppose that "men from Mars" all have two heads and only one foot. You come into a room full of these creatures and decide it is easier to count feet than heads. There are 12 feet in the room. From this you know that there are ____________ men from Mars in the room.

(fill in the number then look at the right answer on the back before reading on)

Whenever you work such an exercise in the lesson, be sure to check the answer on the back immediately afterward. Whether your answer is right or wrong doesn't matter -- the purpose of these exercises is to help you learn. The important thing is to understand why your answer is right or wrong. Each page of the lesson covers only one idea and goes over it 3 or 4 times, so if one paragraph gives you trouble the next paragraph may clear it up for you.

Don't worry about mistakes; they may even help you learn. If you are not sure of an answer, guess or put a "?" in the blank. Work at your own speed, but learn as much as you can in the time you spend. You may write and draw anywhere on the booklet if it helps you to study. Tomorrow we will give you a quiz to see how much you learned from this lesson.

IF YOU HAVE A QUESTION ABOUT THESE INSTRUCTIONS, RAISE YOUR HAND. IF NOT, TURN THE PAGE AND BEGIN THE LESSON.

Experiment IV (February 1962)

Independent Variable

(1) Bypass vs. Short linear method.
(2) Mode of evaluation: Response correctness vs. correctness and understanding as a criterion for bypassing.

Learning methods using the third Sets program:

Short (as defined in Exp. I).
R-Bypass: S was instructed to bypass remedial loops only when his basic step response was correct (the usual method).
RU-Bypass: After each basic step response, S was instructed as follows: "If you understand this page and your answer was right, skip to the next booklet now."

For both Bypass methods the bypassing instruction was repeated on the back of every basic step page, as in the previous (second) Sets program. Each basic step occupied one 7" x 8½" page at the front of the booklet. The remedial loop made up the remaining 3 to 10 pages of each booklet. In the Bypass forms, there was a second opportunity to bypass midway through the loop, in 4 of the longer loops.

Dependent Variables

(1) Score on the 38-item Sets criterion test shown in Appendix A. The items in Sec. V of the test were scored +1 for correct choices and -1 for incorrect choices, so that chance score would be zero.
(2) DAT-Numerical Ability score.
DAT-Verbal Reasoning score.
(3) Time spent in learning.
(4) Number of pages in the entire program read by each S in the Bypass groups, as indicated by responses.

Subjects: Three 8th grade classes (N=58)

Procedure

In each class Ss were assigned by alternation (as in Exp. I) to the 3 learning methods. The program instruction page for the Short and R-Bypass methods was quite similar to those which are shown for the Short and Bypass methods in Exp. V. The RU-Bypass program presented the same instructions as the R-Bypass program except for the bypassing criterion described above. Ss worked continuously until they finished the program. Those who finished early were given an unrelated task as part of a different study. The criterion test was given one day after the program.

Analyses and Results

Analyses of covariance (DAT-Num.) of Sets test scores revealed no significant method differences. Because of previously found ability-by-method interactions, Ss were divided at the 50th percentile on DAT-numerical for comparing the Short and Bypass methods (R and RU combined), using t-tests. Criterion test means were nearly equivalent among the high-ability Ss, while in the low-ability group the Bypass methods slightly surpassed (P<.20) the Short method. The Short group learned in much less time (P<.001). The two Bypass methods did not differ appreciably on either test score or learning time. An item analysis of program responses revealed qualitative differences between the R and RU Bypass method in the kind of procedural errors of bypassing, but the overall number of procedural mistakes was comparable for the two groups.

Table of Means and Estimated Standard Deviations

		Sets test (chance = 0, max. = 33)					
Method	DAT-Numerical	Un-adjusted	DAT-adjusted	High DAT ability	Low DAT ability	Learning Time (Min.)	No. of program pages read
RU-Bypass	62.8	14.1	14.3	16.3	10.2	45.9	51.6
R-Bypass	62.3	13.8	14.0	15.4	11.2	45.8	50.0
Short	65.5	12.1	11.8	15.2	6.14	21.7	
Est. S.D. (within methods)	31.5	7.36	5.34	6.83	5.91	10.9	26.2

Average within-methods correlations	Learning Time	Number of Pages Read	DAT-Num.	DAT-Verbal
Sets test	-.17	-.35	.70	.63
Learning Time		.74	-.39	

Experiment V (March, 1962)

Independent Variables

(1) Bypass vs. linear learning methods.
(2) Order of presentation within each step: basic step preceding remedial loop vs. following loop.

Four learning methods using the third Sets program (See Exp. IV)
Bypass (as defined in Exp. I): The program was identical to that given in the R-Bypass method of Exp. IV.
Short (as defined in Exp. I): The same program as in Exp. IV.
Long (as defined in Exp. I): The full-length third Sets program in normal order.
Long-reversed: The full-length third Sets program, but with remedial loop presented before its basic step.

Dependent Variables

(1) Score on Sets criterion test (Appendix A).
(2) CAT-arithmetic Total Score.
(3) Time spent in learning.

Subjects: Two 4th grade classes (N=47)

Procedure

Ss were assigned to the four methods by alternation (as in Exp. I). The program instruction page for the Bypass group and the three linear groups (Short, Long, Long-reversed) is shown following the results section below. After the first 50 minutes of work, Ss were given a 20-minute recess, as was normal, after which they continued the program until finished. As each S finished, he was given a 3-item questionnaire (questions and answer distributions shown in results section below). Those who finished the program and questionnaire relatively early were given an unrelated learning task as part of another study.

Analyses and Results

An analysis of covariance (CAT-Arith.) of Sets test scores, including all four methods, showed only a marginal method effect (P<.20). The within-methods correlation between CAT and Sets test was only .31. A t-test comparing adjusted test means of the Short and Bypass groups indicated that the Bypass group learned significantly more (P<.05). The Long and Long-reversed groups had nearly identical adjusted test score means and learning times.

Dividing the class into upper and lower halves on CAT score showed an odd pattern of ability-by-method interaction, as can be seen in the table of means below. But the interaction effect was not statistically significant. Learning times were predictably ordered and substantially different among the Short, Bypass and Long groups ($P<.001$).

Table of Means and Estimated Standard Deviations

		Sets test (chance = 0, max. = 33)				
Method	CAT-Arith.	Unadjusted	CAT-adjusted	High CAT ability	Low CAT ability	Learning Time (Min.)
Long	5.48	8.33	8.20	8.17	8.50	86.7
Long-reversed	5.19	7.91	8.30	11.80	4.67	83.3
Bypass	5.53	9.42	9.21	10.67	8.17	57.5
Short	5.42	5.00	4.98	4.83	5.17	26.5
Est. S. D. (within methods)	.82	4.67	4.44	4.67	4.67	14.4

		Frequency distribution for each method				
Question	Answer	Long	Long Reverse	Bypass	Short	Total
1. How hard was this lesson compared to most of your school work?	much harder -	1	0	3	5	9
	a little harder -	6	7	3	5	21
	about the same -	1	2	3	1	7
	a little easier -	0	2	2	1	5
	much easier -	0	0	1	1	2
2. How much of the lesson on sets do you think you understand now?	about all of it	0	2	4	1	7
	most of it	7	6	4	5	22
	some of it	1	3	4	3	11
	just a little of it	0	0	0	4	4
	none of it	0	0	0	0	0
3. Would you like to have more of your school work given in the same way as this lesson?	Yes	1	5	5	6	16
	Can't decide	5	5	5	3	18
	No	2	1	2	5	10

(This was the instruction page of the third Sets program for all learning methods, except that for all linear methods the 4th paragraph was deleted, as noted below.)

INSTRUCTIONS

These booklets are a first lesson on reasoning about sets. Work through them at your own speed, but learn as much as you can in the time you spend. You may write or draw anywhere on the booklet if it helps you to study.

On most pages there are short exercises for you to work with your pencil (a blank to fill in or a question to answer). These are supposed to keep you thinking as you read. But don't get stuck on them. If you can't think of the answer, put a checkmark (✓) in the blank (so we'll know you read that page) and go on.

The right answer to each exercise is on the back of that page. Always compare this with your own to see whether you had the right answer or not. Don't worry about mistakes; they may even help you learn. THIS IS NOT A TEST, so please don't erase your mistakes. (Tomorrow we will give you the test to see how much you learned from this lesson.)

Bypass only

Each booklet in your stack covers only one idea and goes over it several times. We want you to spend the most time on the ideas that you find hardest. So on the backs of certain pages you will be told to skip directly to the next booklet without finishing the booklet you have started, if your answers were right on that page. Skip only as directed on the backs of certain pages.

ANY TIME YOU HAVE A QUESTION ABOUT WHAT YOU SHOULD DO, RAISE YOUR HAND. TURN THE PAGE AND BEGIN THE LESSON.

Experiment VI (March 1962)

Independent Variables

(1) Bypass vs. Short linear form of third Sets program (as described in Exp. V)
(2) Bypass vs. Short linear form of the second United Nations program. These two methods are described in detail on pp. 26-27 of Part One.

Dependent Variables

(1) Ability scores:
    DAT- Numerical Ability
    DAT - Verbal Reasoning
(2) Sets program variables:
    Criterion test (Appendix A)
    Learning time
(3) U. N. program variables:
    Criterion test (42 items, about half completion and half multiple-choice).
    Learning time.

Subjects: 54 students from two 12th grade classes (government and English).

Procedure

All Ss began with the Sets program (instructions shown on the previous page), having been "randomly" assigned the Short or Bypass method. The Bypass group worked continuously to completion of the Sets program and then resumed their normal schoolwork.

All Ss given the Short Sets program were given the U. N. program immediately afterward, half getting the Short form and half the Bypass form of the U. N. program on a random basis. The half page of instructions for the Bypass U. N. program, and for the Short form, are shown following the results section below. All Ss taking the U. N. program were given the U. N. criterion test immediately after completing the program. There were 110 minutes available for this experimental sequence and all Ss but one finished easily in that interval. On the following day at the same hour all Ss were given the Sets criterion test.

Analyses and Results

Sets Program and Test

The Sets criterion test correlated .71 with DAT-Numerical and .59 with DAT-Verbal. Unadjusted test means for the Short and Bypass methods were very close and when adjusted by DAT_Numerical score, the means were nearly identical. When the sample was divided at the DAT-Numerical 50th percentile, the high ability group scored higher by both methods, but no ability-by-method interaction was evident. Learning times were nearly twice as long for the Bypass group as for the Short group (P<.001).

Table of Means and Estimated Standard Deviations

Method	DAT-Numerical	Sets test (chance = 0, max. = 33) Unadjusted	DAT-Num. Adjusted	High DAT ability	Low DAT ability	Learning Time (Min.)
Bypass	71.1	18.3	17.6	21.6	8.3	54.8
Short	65.2	17.1	17.6	20.2	11.7	29.9
Est. S.D. (within methods)	27.5	7.22	5.0			13.9

U. N. Program and Test

The U. N. criterion test correlated .43 with DAT-Verbal and .09 with DAT-Numerical. The Bypass group test mean significantly exceeded the Short group mean (analysis of variance, P<.05). Splitting the sample at the 60th percentile on DAT-Verbal, the high ability Ss scored slightly higher (not significantly) than the low ability group but there was no sign of an ability-by-method interaction. A 12th grade control group from a different school (Exp. VIII) was given the U. N. criterion test with no program preceding. An analysis of covariance (DAT-Verbal) of test scores of the Short, Bypass and control groups (see table of means) yielded an F significant at the .001 level. Mean learning time for the Bypass group was 3 times as large as the Short group mean with no overlap between the distributions.

Table of Means and Standard Deviations

Method	DAT-Verbal	U.N. test (chance=7.75, max.=43.8) Unadjusted	DAT-Adjusted	High DAT ability	Low DAT ability	Learning Time (Min.)
Bypass	67.1	32.2	32.9	39.3	37.5	45.9
Short	65.0	38.4	38.8	33.5	30.3	15.8
Control*	73.6	25.1	24.5			00.0
Est. S. D.	24.6	6.9	5.9			

* From a different school (Exp. VIII)

BYPASS U.N. Program (Second version)

Instructions

This booklet is to help you learn about the United Nations. It will give you information and then ask you questions about what you have read. Some of the questions will be multiple-choice. Answer them by putting a circle around the letter in front of the correct answer, like this:

On sunny days the sky is
A gray
[✓ B] blue
C red
D green

Sometimes there will be a sentence with a blank in it, and you will fill in the blank in such a way as to make a true statement, like this:

On sunny days the color of the sky is blue.

Answer all the questions from memory; DO NOT LOOK BACK TO FIND THE ANSWERS. Sometimes, if you answer the questions right and show that you have learned the material, the book will tell you to skip ahead to a page farther on. Unless the book tells you to skip, just continue reading.

Now turn to page 1 and begin.

SHORT U.N. Program (Second version)

Instructions

This booklet is to help you learn about the United Nations.

Read it through once, carefully, and try to remember what you have read.

Now turn to page 1 and begin.

Experiment VII (March 1962)

Independent Variable

Order of presentation within each step of third Sets program: basic step preceding remedial loop vs. following remedial loop. The two learning methods compared were the Long and Long-reversed methods as defined in Exp. V.

Dependent Variables

(1) Sets criterion test (Appendix A)
(2) Ability scores:
    CTMM - Total Mental
    CTMM - Language
    CTMM - Non-language
(3) Learning time
(4) Previous study of set theory

Subjects: Two 8th grade classes (N = 47)

Procedure

Ss were assigned randomly to the two learning methods (except that 5 Ss were reserved for a pilot study of retention from the U.N. program--results are reported under Exp. VIII). Except for one 2-minute break, Ss worked continuously on the Sets program until finished (see Exp. V for the instruction page of both programs.) The criterion test was given one day later.

Analyses and Results

Correlations between the Sets test and CTMM scores were:

CTMM - Total Mental	.49
CTMM - Language	.48
CTMM - Non-language	.39

Thirteen Ss had taken a math course the previous year which included the topic of set theory. These Ss scored much higher ($P < .001$) on the Sets test than those who had not taken the math course, and the difference was equivalent for the two methods. No analysis was performed to compare the two methods among the "naive" Ss, since test means differed only slightly and in the same direction as the small difference in CTMM means (see table below). Learning times did not differ appreciably for the two methods.

Table of Means and Standard Deviations

Method	CTMM Total	Sets test (0-33) No previous course	Sets test (0-33) Previous course	Learning Time (Min.)
Long	75.1	15.2	26.0	81.9
Long-reversed	72.6	15.6	25.0	84.2
S. D.		6.4	3.8	

Experiment VIII (April 1962)

Independent Variables

(1) Bypassing vs. linear methods of learning:
    (a) the third Sets program (see Part One for description). Bypass, Long and Short methods as defined in Exp. I.
    (b) the second United Nations program. Bypass vs. Short method (the same as in Exp. VI)

(2) Grade level
    Sets program: grades 4, 6, 8, 10 and 12
    U.N. program: grade 8

Dependent Variables

(1) Ability scores:
    CAT - Arithmetic Total (grades 4 and 6)
    DAT - Numerical Ability (grades 8, 10 and 12)
    DAT - Verbal Reasoning (grade 8, U.N. class only)

(2) Sets program variables:
    Criterion test (Appendix A)
    Program score--number of correct responses to the 15 basic steps (no loop responses included).
    Learning time (minutes spent on the program)

(3) U.N. program variables:
    Criterion test (same as in Exp. VI)
    Learning time

Subjects: (total N = 780)

4th grade: 3 classes from each of two schools (N = 166)
6th grade: 3 classes from each of two schools (N = 173)
(The 4th and 6th grades were from the same 2 schools, and all students at these grade levels served as Ss)
8th grade: 6 classes from a third school, each class being a heterogeneous, essentially random sample of the 8th grade at that school. (N = 150)
10th grade: Two heterogeneous physical education classes of boys and two of girls (N = 150)
12th grade: Three heterogeneous physical education classes of boys and three of girls (N = 139)
(The 10th and 12th grades were sampled from the same high school)

Procedure

In addition to the regular research staff, a number of graduate and undergraduate students from Stanford University served as experimenters during data-collection. One 8th grade class was chosen at random to be given the U.N. program. Ss were assigned by alternation (see Exp. I) to the Short or Bypass method (see Exp. VI for instruction pages) and given as much time as they needed to finish the U.N. program. When all Ss had finished the program, the class was given a 10-minute recess, then all were given the U.N. criterion test. In one other 8th grade class, all Ss in the Short Sets program group were given the U.N. criterion test without the U.N. program and thus served as a no-learning control group. In five 8th grade classes and in all 4th and 6th grade classes, Ss were assigned the Bypass, Long or Short learning methods by alternation and given the corresponding Sets program. The instruction page for all methods is shown under Exp. V (p. 47). Recesses were taken at normal times.

The 10th and 12th grade classes were assembled by pairs, one boys' class and one girls' class from the same grade in the same room. Ss were given the Bypass or Short form of the Sets program (by alternation) and worked until they had completed it or until the 50 min. period ended (see discussion of time limits in Part One, p. 22). One-third of the 12th grade Ss (all in one room) were kept especially quiet throughout the learning. The results of comparing this group with the other 12th grade groups (no differences) are discussed in Part One, p. 23. The Short group in this "quiet" class also served as the U.N. control group in Exp. VI.

All Ss who took the Sets program were given the Sets criterion test 24 hours later. Nearly all Ss spent between 10 and 20 minutes on the test.

Analyses and Results

Sets Program and Test

A preliminary analysis of covariance (CAT-Arith.) of Sets test scores in the 6th grade yielded a method-by-class interaction of small magnitude ($P < .05$). Since class differences appeared largest by inspection in this grade, we decided to discontinue use of "class" as a nested variable (within grades) in order not to cross-classify the data to death. Instead, "mathematical ability" was treated as a "levels" variable across all grades. The entire sample of Ss at a given grade level was divided into three ability groups (high, medium, and low) of about equal size. Since Ns had been equalized between methods for each class rather than for each ability group, the resulting 3-factor (learning method, grade, and ability group) factorial design for analysis of variance had fairly disproportionate Ns. After correcting Ns to proportionality, analyses of variance of both Sets test scores and learning times were conducted using this design. Since only two learning methods were used in the high school, two overlapping analyses were performed for each dependent variable (learning time and test score). Both analyses included all 3 ability groups, but one analysis compared the Short and Bypass methods across all 5 grade levels while the second analysis compared all 3 methods across the 4th, 6th and 8th grades. The following pages show the summaries of the analyses of variance, and tables of means, first for Sets test scores and then for learning times.

Analysis of Sets test scores

(grades 4 to 12; 3 ability levels; Short vs. Bypass)

Source of Variance	df	SS	MS	F	P-level of signif.
Total	518	40,321			
Grade (G)	4	14,519	3,630	110.5	.001
Ability (A)	2	7,996	3,998	121.7	.001
Method (M)	1	175	175	5.3	.025
G x A	8	1,179	147	4.5	.001
G x M	4	310	77	2.4	.10
A x M	2	0	0	-	N.S.
G x A x M	8	85	10	-	N.S.
Within	489	16,057	33		

Analysis of Sets test scores

(grades 4 to 8; 3 ability levels; 3 methods)

Source of Variance	df	SS	MS	F	P-level of signif.
Total	453	22,385	49		
Grade (G)	2	7,428	3,714	166.6	.001
Ability (A)	2	3,716	1,858	83.4	.001
Method (M)	2	747	373	16.8	.001
G x A	4	623	156	7.0	.001
G x M	4	332	83	3.7	.005
A x M	4	0	0	-	N.S.
G x A x M	8	23	3	-	N.S.
Within	427	9,516	22		

Table of Sets test means for High, Medium, and Low Ability Groups, and for All Abilities Combined (Total)

(Est. S. D. within-methods is approximately 5.0)

		Learning Method		
		Short	Bypass	Long
12th grade	Total	18.6	18.6	
	High	27.1	26.3	
	Med.	17.9	16.9	
	Low	11.8	12.9	
10th grade	Total	14.5	16.4	
	High	20.6	23.7	
	Med.	11.5	12.1	
	Low	10.0	11.1	
8th grade	Total	11.7	15.7	17.2
	High	17.9	20.7	24.3
	Med.	10.3	14.6	15.7
	Low	7.5	11.8	11.5
6th grade	Total	6.2	7.2	9.2
	High	9.4	10.1	12.0
	Med.	5.4	7.6	9.5
	Low	4.3	3.8	6.0
4th grade	Total	4.8	4.7	6.0
	High	7.8	6.5	7.9
	Med.	4.7	5.3	6.7
	Low	2.1	1.9	3.2

Analysis of Sets learning times

(grades 4 to 12; 3 ability levels; Short and Bypass)

Source of Variance	df	SS	MS	F	P-level of signif.
Total	518	194,249			
Grade (G)	4	42,911	10,728	68.4	.001
Ability (A)	2	2,792	1,396	8.9	.001
Method (M)	1	57,036	57,036	363.5	.001
G x A	8	726	91	-	N.S.
G x M	4	11,651	2,913	18.6	.001
A x M	2	1,219	609	3.9	.025
G x A x M	8	1,180	147	-	N.S.
Within	489	76,734	157		

Analysis of Sets learning times

(grades 4 to 8; 3 ability levels; 3 methods)

Source of Variance	df	SS	MS	F	P-level of signif.
Total	453	290,457			
Grade (G)	2	13,747	6,873	26.5	.001
Ability (A)	2	515	257	-	N.S.
Method (M)	2	156,603	78,301	301.4	.001
G x A	4	1,920	480	1.9	.20
G x M	4	2,120	530	2.0	.10
A x M	4	1,898	474	1.8	.20
G x A x M	8	2,716	339	1.3	N.S.
Within	427	110,938	260		

Table of Sets learning time means for High, Medium and Low Ability Groups, and for All Abilities (Total)

		Learning Method		
		Short	Bypass	Long
12th grade	Total	24	35	
	High	23	30	
	Med.	23	33	
	Low	26	39	
10th grade	Total	24	32	
	High	23	29	
	Med.	25	36	
	Low	26	34	
8th grade	Total	24	48	71
	High	24	38	66
	Med.	25	46	76
	Low	25	57	71
6th grade	Total	28	59	76
	High	27	60	82
	Med.	27	60	68
	Low	30	58	77
4th grade	Total	36	67	78
	High	35	62	78
	Med.	35	67	79
	Low	38	74	77

In the following correlation tables, each entry is based on an N of 45 to 75 Ss.

Correlations: Sets Program Score with Learning Time	Learning Method Short	Bypass
Grade 12	-.39	-.56
Grade 10	-.22	-.47
Grade 8	-.12	-.62
Grade 6	-.21	-.14
Grade 4	.01	-.05

Correlations: Sets Test with Sets Program Score	Learning Method Short	Bypass
Grade 12	.59	.37
Grade 10	.53	.50
Grade 8	.55	.43
Grade 6	.05	.15
Grade 4	.29	.12

Correlations: Sets Test with Learning Time	Learning Method Short	Bypass	Long
Grade 12	-.22	-.33	
Grade 10	-.13	-.17	
Grade 8	-.11	-.39	.26
Grade 6	.10	.30	.09
Grade 4	-.33	.09	.36

Correlations: Sets Test with Mathematical Ability	Learning Method Short	Bypass	Long
Grade 12 (DAT)	.70	.63	
Grade 10 (DAT)	.50	.67	
Grade 8 (DAT)	.53	.61	.41
Grade 6 (CAT)	.55	.55	.55
Grade 4 (CAT)	.66	.55	.44

U.N. Program and Test (Analyses and Results)

The correlations between U.N. test score and DAT-Verbal score were:

Bypass method	.77
Short method	.63
Control	.35

An analysis of covariance (DAT-Verbal) showed highly significant differences among the three method groups ($P < .001$). The Bypass group, as in Exp. VI, spent 3 times as long on the program as the Short group, with no overlap between the distributions.

Table of Means and Estimated Standard Deviations

		U.N. test (chance = 7.75, max. = 43.8)		Learning Time (Min.)
Method	DAT Verbal	Unadjus.	DAT-adjus.	
Bypass	62.9	33.1	34.5	48.8
Short	60.6	29.5	31.0	16.4
Control	88.0	25.6	21.6	0.0
Est. S. D. (within methods)	Control: 9.4 Others: 28.6	8.0	5.3	

Five 8th grade students from another school (Exp. VII) provided a tentative indication that the learning achieved with the U.N. program was not quickly forgotten. All 5 were given the U.N. Bypass program followed immediately by the test. One day later they were given the same U.N. test with no intervening feedback on previous test performance. Mean test score on the first day was 40.2, and on the second day 40.5.

REFERENCES

Briggs, L. J., Goldbeck, R. A., Campbell, V. N., and Nichols, D. G. Experimental results concerning form of response, size of steps, and individual differences in automated programs. In *Proceedings of the Symposium on the Application of Digital Computers to Automated Instruction*. New York: Wiley, 1962, in press.

Campbell, V. N. Adjusting self-instruction programs to individual differences: studies of cueing, responding and bypassing. San Mateo: American Institute for Research. AIR-C41-7/61-SR, 1961.

Campbell, V. N., Terry, D. F., and Shearer, J. W. Some basic concepts of set theory. San Mateo: American Institute for Research. AIR-C41-4/62-RP, 1962.

Curr, W., and Gaulay, N. An experimental evaluation of remedial education. *Brit. J. Educ. Psychol.*, 23:45-57, February, 1953.

Gagne, R. M. Military training and principles of learning. *Amer. Psychologist*, 1962, *17*, 83-91.

Goldbeck, R. A. and Campbell, V. N. The effects of response mode and response difficulty on programed learning. *J. educ. Psychol.*, 1962, in press.

Green, B. Attitude measurement. In G. Lindzey (ed.), *Handbook of Social Psychology*. Cambridge: Addison-Wesley, 1954.

Haslerud, G. M., and Meyers, S. The transfer value of given and individually derived principles. *J. educ. Psychol.*, 1958, *49*, 293-298.

Hendrix, G. A new clue to transfer of training. *Elem. sch. J.*, 1946, *48*, 197-208.

Lumsdaine, A. A. and Glaser, R. (eds.) *Teaching machines and programmed learning*. Washington, D.C.: National Education Association, 1960.

Valentini, H. B. Some results of remedial education in a child guidance centre. *Brit. J. Educ. Psychol.*, 21:145-9, June , 1951.

Appendix A

Sets criterion test
used in conjunction
with the third Sets
program

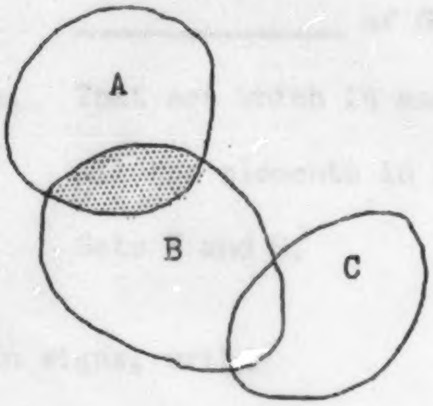
Name ____________________

This is a quiz to see how much you learned about sets. Work rapidly. Skip questions you can't answer and come back to them later if there is time. Guess if you are still not sure of an answer. You may write or draw anywhere on this booklet.

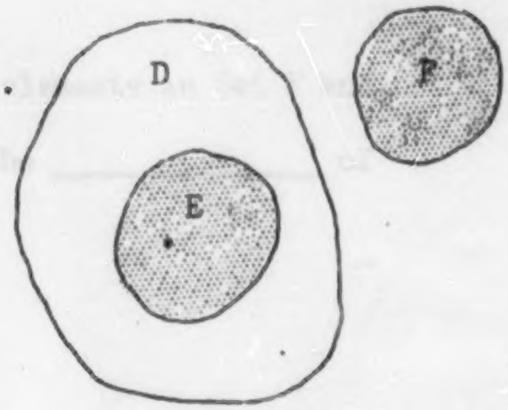
TURN THE PAGE AND BEGIN.

I. At the left of each picture, write in signs what the shaded part of the picture is. The first two answers are given as examples. Notice that in each picture one of the sets is not needed to describe the shaded area. Use only letters and these two signs where needed: $\sqcap$, $\sqcup$ (and parentheses where needed).

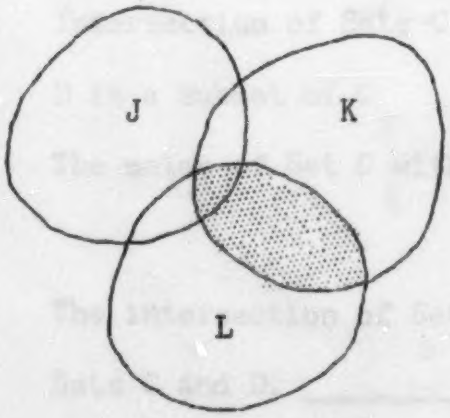
$A \sqcap B$



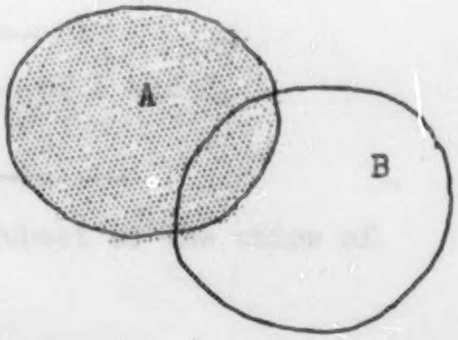
$E \sqcup F$



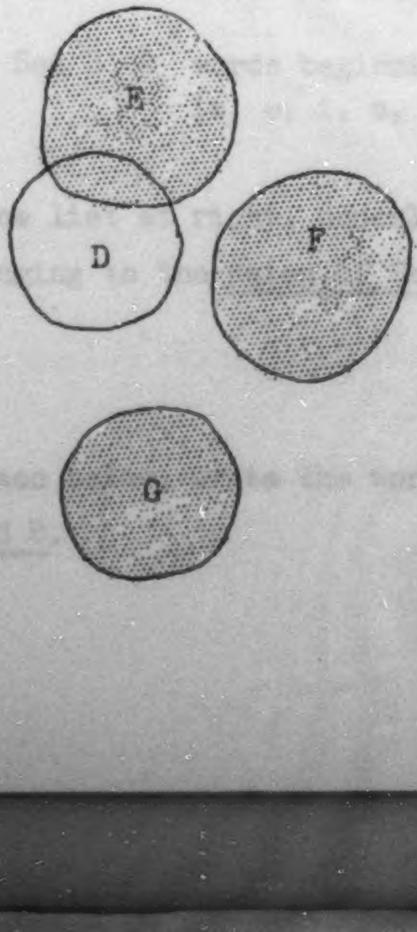
____



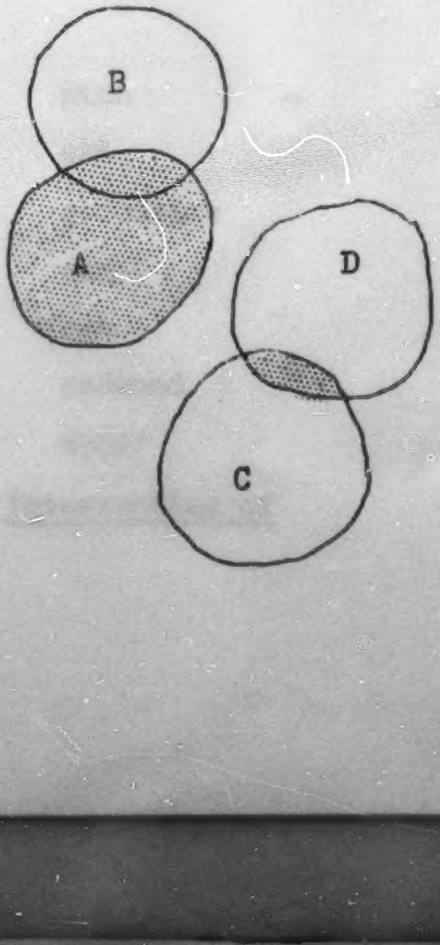
____



____



____



II. Definitions (Fill in the blanks).

a. If every element of Set A is also an element of Set B, then A is a ____________ of B.

b. The part of Set A which is also in Set B is called the ____________ of Sets A and B.

c. That set which is made up of all the elements in Set F and all the elements in Set G is called the ____________ of Sets F and G.

III. In signs, write

a. Union of Sets C and D ____________

b. Intersection of Sets C and D ____________

c. D is a subset of C ____________

d. The union of Set C with the intersection of D and E.

____

e. The intersection of Sets C and D is a subset of the union of Sets C and D. ____________

IV. In the list of words shown at right,

Set A = names of trees

Set B = words beginning with vowels (a, e, i, o, u.)

In the list at right, underline all words belonging to the union of Sets A and B.

pine
eat
groceries
elm
oak
redwood
upper

In the space below, write the words belonging to the intersection of Sets A and B.

V. Below are several statements in signs about the picture at the right. Circle the true statements.

$D \sqsubset C$

$C \sqsubset D$

$B \sqsubset C$

$D \sqsubset K$

$A \sqsubset B$

$(A \sqcup B) \sqsubset K$

$(B \sqcup D) \sqsubset C$

$(B \sqcap C) \sqsubset B$

$(B \sqcap C) \sqsubset (B \sqcup C)$

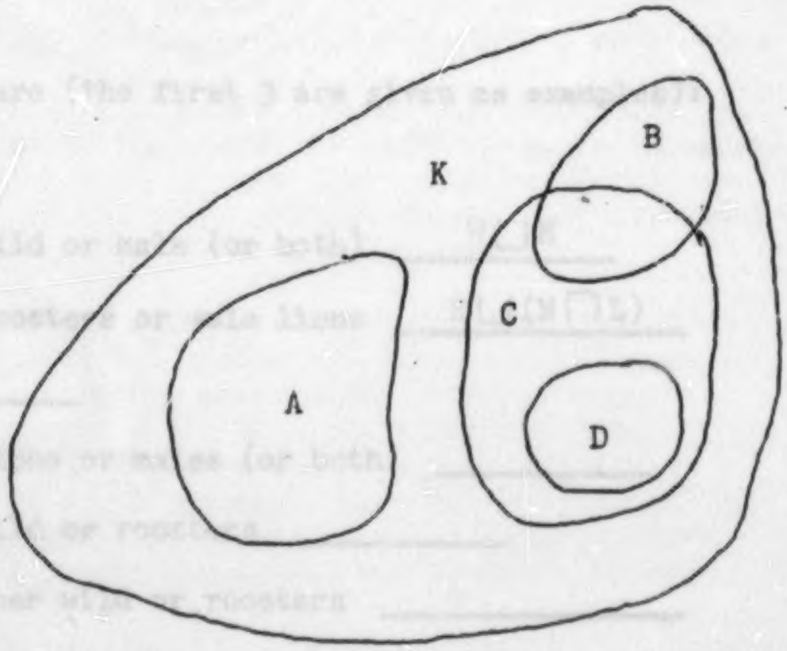
$A \sqsubset (B \sqcap K)$

$A \sqsubset (B \sqcup K)$

$C \sqsubset (A \sqcup B \sqcup D)$

$(B \sqcap C \sqcap K) \sqsubset D$

$(B \sqcap C \sqcap K) \sqsubset (C \sqcup A)$



VI.

Set W = Wild animals

Set M = Male animals

Set L = Lions

Set R = Roosters

Write in signs what the following sets are (the first 3 are given as examples):

1. all male lions $M \sqcap L$
2. all animals which are either wild or male (or both) $W \sqcup M$
3. all animals which are either roosters or male lions $R \sqcup (M \sqcap L)$
4. all male wild animals __________
5. all animals which are either lions or males (or both) __________
6. all animals which are either wild or roosters __________
7. all male animals which are either wild or roosters __________
8. all wild male lions __________
9. all animals which are either roosters or male lions __________

Write the following statements in signs (the first 2 are given as examples):

10. All roosters are male animals. $R \sqsubset M$
11. All male wild animals are wild. $(M \sqcap W) \sqsubset W$
12. All lions are wild animals. __________
13. All male lions are wild animals. __________
14. All lions are either male or wild (or both). __________
15. All male lions are either wild animals or roosters. __________

Appendix B

Index of Difficulty Interdependence

An index, D, was developed as a measure of the extent to which a set of test items or program responses have the same rank order of difficulty for a given set of persons. The procedure for computing D is as follows:

(1) Record as right or wrong the answer of each person to each item.

(2) Put this record in the form of a table or matrix (N persons by K items) with the items in rank-order of difficulty, as measured by the number of persons missing each item. (Items of the same difficulty may be put in any order without appreciably affecting the index.) The order of persons does not matter.

(3) For each person, divide the column (or row) of items into two segments so as to maximize the number of right answers in one segment and the number of wrong answers in the other segment. For most persons there will still be a few misplaced items, i.e., right answers in the "wrong answer" segment and wrong answers in the "right answer" segment. Call these "errors." This step of the task, then, is to divide the column of items at that point which minimizes the number of "errors" for each person.

(4) Count "E" the number of "errors" for all persons summed together.

(5) For each person count the number of right answers or the number of wrong answers, whichever is smaller. Sum the smaller of these two numbers for all persons together and call the sum "T." "T" is the number of "errors" you would make by predicting that each person's answers were all right if his score was over 50%, or all wrong if his score was under 50%.

(6) Compute $D = 1 - \frac{E}{T}$. If all items were independent in their difficulties, E would approach T as a maximum value, and D would be close to zero. Conversely, D approaches 1.00 as E approaches 0, and D = 1 would mean that the order of item difficulties was the same for all persons.